

March 1990

DTIC FILE COPY

UILU-ENG-90-2208
DAC-18

②

COORDINATED SCIENCE LABORATORY
College of Engineering

AD-A219 654

**STATISTICAL DESIGN
OF MOS VLSI
CIRCUITS WITH
DESIGNED
EXPERIMENTS**

Tat-Kwan Edgar Yu

DTIC
S **ELECTE** **D**
MAR 23 1990
E

UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Approved for Public Release. Distribution Unlimited.

90 03 23 004

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

| | | | | |
|---|-------|---|--|----------------------------|
| 1a. REPORT SECURITY CLASSIFICATION Unclassified | | | 1b. RESTRICTIVE MARKINGS None | |
| 2a. SECURITY CLASSIFICATION AUTHORITY | | | 3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited | |
| 2b. DECLASSIFICATION/DOWNGRADING SCHEDULE | | | | |
| 4. PERFORMING ORGANIZATION REPORT NUMBER(S) UILU-ENG-90-2208 (DAC-18) | | | 5. MONITORING ORGANIZATION REPORT NUMBER(S) | |
| 6a. NAME OF PERFORMING ORGANIZATION Coordinated Science Lab University of Illinois | | 6b. OFFICE SYMBOL (If applicable) N/A | 7a. NAME OF MONITORING ORGANIZATION Office of Naval Research | |
| 6c. ADDRESS (City, State, and ZIP Code) 1101 W. Springfield Ave. Urbana, IL 61801 | | | 7b. ADDRESS (City, State, and ZIP Code) Arlington, VA 22217 Research Triangle Park, NC 27709 | |
| 8a. NAME OF FUNDING/SPONSORING ORGANIZATION Joint Services Electronics Program & SRC | | 8b. OFFICE SYMBOL (If applicable) | 9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER N00014-84-C-0149 SRC 89-DP-Antic-Trick | |
| 8c. ADDRESS (City, State, and ZIP Code) Arlington, VA 22217 | | | 10. SOURCE OF FUNDING NUMBERS | |
| | | | PROGRAM ELEMENT NO. | PROJECT NO. |
| | | | TASK NO. | WORK UNIT ACCESSION NO. |
| 11. TITLE (Include Security Classification) Statistical Design of MOS VLSI Circuits with Designed Experiments | | | | |
| 12. PERSONAL AUTHOR(S) Yu, Tat-Kwan Edgar | | | | |
| 13a. TYPE OF REPORT Technical | | 13b. TIME COVERED FROM _____ TO _____ | 14. DATE OF REPORT (Year, Month, Day) 1990 March | 15. PAGE COUNT 112 |
| 16. SUPPLEMENTARY NOTATION | | | | |
| 17. COSATI CODES | | | 18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number) | |
| FIELD | GROUP | SUB-GROUP | Statistical design; VLSI design; Experimental design, | |
| | | | Thesis | |
| 19. ABSTRACT (Continue on reverse if necessary and identify by block number) | | | | |
| <p>A new approach for the statistical design and analysis of MOS is introduced. The proposed approach approximates the circuit performances, such as gain and delay, by fitted models. The fitted models are then used as surrogates of the circuit simulator to predict and optimize the parametric yield with computation efficiency and to achieve off-line quality control. The use of statistical design and analysis of experiments for model construction have been investigated theoretically and experimentally, and different methods to assess the adequacy of a fitted performance model have been studied.</p> | | | | |
| 20. DISTRIBUTION/AVAILABILITY OF ABSTRACT <input checked="" type="checkbox"/> UNCLASSIFIED UNLIMITED <input type="checkbox"/> SAME AS RPT. <input type="checkbox"/> DTIC USERS | | | 21. ABSTRACT SECURITY CLASSIFICATION Unclassified | |
| 22a. NAME OF RESPONSIBLE INDIVIDUAL | | | 22b. TELEPHONE (Include Area Code) | 22c. OFFICE SYMBOL |

STATISTICAL DESIGN OF MOS VLSI CIRCUITS WITH DESIGNED EXPERIMENTS

BY

TAT-KWAN EDGAR YU

B.S.E., University of Michigan, 1982
M.S., University of Illinois, 1985

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Electrical Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 1989

Urbana, Illinois

| | |
|-------------------|-------------------------------------|
| Accession For | |
| NTIS GRA&I | <input checked="" type="checkbox"/> |
| DTIC TAB | <input type="checkbox"/> |
| Unannounced | <input type="checkbox"/> |
| Justification | |
| By | |
| Distribution/ | |
| Availability Code | |
| Dist | Avail and/or Special |
| A-1 | |



UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

THE GRADUATE COLLEGE

JULY 1989

WE HEREBY RECOMMEND THAT THE THESIS BY

TAT-KWAN EDGAR YU

ENTITLED STATISTICAL DESIGN OF MOS VLSI CIRCUITS

WITH DESIGNED EXPERIMENTS

BE ACCEPTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR

THE DEGREE OF DOCTOR OF PHILOSOPHY

S. M. Kang / T. N. Tzeng

Director of Thesis Research

Timothy N. Tzeng

Head of Department

Committee on Final Examination†

S. M. Kang

Chairperson

Timothy N. Tzeng

Wachin Huij

James Scher

† Required for doctor's degree but not for master's.

STATISTICAL DESIGN OF MOS VLSI CIRCUITS WITH DESIGNED EXPERIMENTS

Tat-Kwan Edgar Yu, Ph.D.
Department of Electrical and Computer Engineering
University of Illinois at Urbana-Champaign, 1989
S. M. Kang, Advisor

Many methods for the statistical design and analysis of integrated circuits have been proposed over the past years. However, these methods typically require a large number of computationally expensive circuit simulator runs, and their applications are limited to small circuits.

This research investigated new approaches for the statistical design and analysis of MOS integrated circuits. This work has resulted in a new and efficient circuit performance modeling approach to statistical design. The proposed approach approximates the circuit performances, such as gain and delay, by *fitted models of the inputs to the circuit simulator*. The computationally inexpensive fitted models are then used as surrogates of the circuit simulator to predict and optimize the parametric yield and to achieve off-line quality control. The use of statistical design and analysis of experiments for model construction have been investigated theoretically and experimentally, and different methods to assess the adequacy of a fitted performance model have been studied.

ACKNOWLEDGEMENTS

I would like to express my appreciation and gratitude to Professor Sung-Mo Kang, my dissertation advisor, for his invaluable guidance throughout the course of this research. With his great personality, he provided support and encouragement in all respects of my life throughout my graduate study. I would also like to thank my co-advisors Professors Ibrahim N. Hajj and Timothy N. Trick for their guidance and encouragement.

I am grateful to Professors William J. Welch and Jerome Sacks for their advice and valuable discussions on statistical design and analysis techniques.

I also thank Dave Smart, Su Shun Lin, Seung Hoon Lee, and other members of the Circuits and Systems Group at the Coordinated Science Laboratory, University of Illinois, for many interesting discussions.

Finally, I would like to thank my mother, Edna, and a very personal friend, Ng Kam Yau, for their love and support.

The financial support of the Semiconductor Research Corporation, the Joint Services Electronics Program, and the AT&T Bell Laboratories is acknowledged. The use of word processing facilities at Motorola Inc. for the final preparation of this thesis is also acknowledged.

TABLE OF CONTENTS

| | |
|--|----|
| 1. INTRODUCTION | 1 |
| 2. STATISTICAL MODELING OF THE MOS TRANSISTOR | 5 |
| 2.1. Best- and Worst-case Analyses | 6 |
| 2.2. Critical MOSFET Parameters | 7 |
| 2.3. Discussion | 10 |
| 3. PARAMETRIC YIELD PREDICTION | 11 |
| 3.1. Previous Parametric Yield Prediction Methods | 12 |
| 3.1.1. Monte Carlo methods | 13 |
| 3.1.2. Statistical performance modeling method | 15 |
| 3.2. Uncertainty Analysis | 17 |
| 3.3. Proposed Parametric Yield Prediction Method | 19 |
| 3.4. Yield Prediction Examples | 20 |
| 3.5. Concluding Remarks | 35 |
| 4. PARAMETRIC YIELD OPTIMIZATION | 38 |
| 4.1. Proposed Parametric Yield Maximization Method | 39 |
| 4.2. Application to Yield Maximization of CMOS Analog Circuits | 42 |
| 4.3. Concluding Remarks | 59 |
| 5. PARAMETER DESIGN METHOD FOR OFF-LINE QUALITY CONTROL | 60 |
| 5.1. Taguchi's Parameter Design Method | 61 |
| 5.2. Proposed Off-line Quality Control Method | 63 |
| 5.3. Example: Minimization of Process Dependent Clock Skew | 64 |
| 5.3.1. Modeling the circuit performances | 67 |
| 5.3.2. Modeling the loss statistics directly | 72 |
| 5.3.3. Results and comparisons | 75 |
| 5.4. Discussion | 78 |
| 6. CONCLUSIONS | 80 |

| | |
|--|-----|
| APPENDIX A. THE DESIGN OF EXPERIMENT | 84 |
| A.1. Introduction | 84 |
| A.2. Integrated Mean-squared Error Criterion | 85 |
| A.3. Example of Design Point Selection by ACED | 89 |
| A.3.1. All-variance design | 89 |
| A.3.2. All-bias design | 89 |
| A.3.3. Selection of a robust design | 92 |
| A.4. Design and Analysis of Computer Experiments | 93 |
| A.5. Latin Hypercube Design | 94 |
| APPENDIX B. MODEL ASSESSMENT | 97 |
| B.1. A Statistical F-test Procedure for Model Assessment | 97 |
| B.2. Assessing the Goodness of Fit | 101 |
| REFERENCES | 102 |
| VITA | 106 |

CHAPTER 1.

INTRODUCTION

During the past decade, the feature sizes of the Metal Oxide Semiconductor (MOS) transistors in Very Large Scale Integrated (VLSI) circuits have been scaled down rapidly, and this trend is expected to continue into the 1990's. Despite the technological progress in patterning the features of the MOS transistors, the statistical variations in the MOSFET parameters, such as the channel threshold voltage and gate oxide thickness, have not been scaled down in proportion. As a result, the circuit performances become even more sensitive to these uncontrollable statistical variations. To ensure the manufacturability of a circuit, statistical variabilities must be considered in the design procedure.

A problem in the statistical design of MOS integrated circuits is the modeling of the device parameter distribution. Traditionally, circuit simulations at the "best-" and "worst-case" process files have been used to evaluate the "range" of the circuit performance variations. Best- and worst-case circuit simulations do not estimate the distributions of the circuit performances, however. In Chapter 2 we review the best- and worst-case approach, and present an improved model of the MOSFET parameters [1]. The proposed model assumes the statistical variations in the circuit performances are mainly due to a small subset of critical parameters. It models the non-critical device parameters as functions of the critical parameters. This model of the MOSFET parameters [1], incorporated into our circuit design algorithms, significantly reduces the complexity of the statistical analysis.

The critical bottleneck in statistical circuit design is the high cost of the circuit simulations. Existing design methods typically require a large number of runs. In this thesis we present a new design approach which significantly reduces the simulation cost. Our approach assumes that the circuit performances, such as gain or delay, can be approximated by computationally-inexpensive functions of the inputs to the circuit simulator. These inputs are the designable circuit parameters, the operating conditions, and the distribution of the device parameters. We fit the functions to data collected from a *statistically designed experiment* incurring relatively few runs of the circuit simulator. These fitted models act as computationally inexpensive surrogates of the circuit simulator for performance prediction. The applications of our design approach to parametric yield prediction, yield optimization, and off-line quality control are presented in Chapters 3, 4, and 5, respectively.

Circuit designers often measure a circuit's quality by *parametric yield*, which is the percentage of the functionally good chips that satisfy the constraints. Parametric yield prediction is important because it assesses the profitability of a design, before a large amount of resources is invested in the manufacturing. In Chapter 3 we examine the existing yield prediction methods and some related research in statistics and then present our improved yield prediction algorithm [2,3]. (Unless stated otherwise, the term yield refers to the parametric yield in the rest of the thesis. Other causes of yield loss are explained in [4]). The yield prediction methods in [1,5] model each performance by an approximating function of the uncontrollable statistical variations, and then the Monte Carlo method is used with the fitted models to predict yield. However, the data are collected by empirical methods in [1,5], and no attempt has been made to assess the prediction capabilities of the fitted models. In the proposed method, we fit the performance models to data generated according to a statistically

designed experiment, and introduce systematic procedures to assess the prediction capabilities of the fitted models. Monte Carlo simulation with the fitted models leads to an estimate of the parametric yield. We illustrate through circuit examples the experimental design and model assessment procedures.

Parametric yield optimization is an extension of yield prediction. For a known distribution of the input device parameters, the problem is to maximize the parametric yield with respect to the designable circuit parameters, such as the MOSFET channel widths and aspect ratios. Gradient methods for yield maximization [6] compute yield gradients with respect to the designable circuit parameters, and then use steepest ascent to optimize the parametric yield. Data from many runs of the circuit simulator are needed to compute a single yield gradient, and many gradient iterations are required. As a result, yield gradient methods typically require a very large number of runs, and their applications are limited to small circuits.

In Chapter 4 we extend the method in Chapter 3 to parametric yield optimization [7,8]. We model each circuit performance as a function of all parameters of interest: the designable circuit parameters, the statistical variations, and the operating conditions. Data generated according to a *single* experimental design for all parameters are used to identify and fit the model. For fixed values of the designable parameters, the Monte Carlo yield is estimated with the fitted models. The estimated yield is numerically optimized. We give circuit examples where sufficiently accurate yield estimates and good actual circuit designs can be achieved with about 100 circuit simulator runs.

Taguchi's parameter design method for off-line quality control [9] has generated great interest in the engineering and statistics literature. Instead of maximizing the parametric yield, Taguchi [9] uses statistical design and analysis of experiments to design products

(circuits) that are insensitive to variations in the manufacturing process and/or the environmental conditions. Like the yield gradient methods, Taguchi's design method typically requires many (circuit simulator) runs. Data from many runs are collapsed to estimate a single "signal-to-noise ratios." just as the yield gradient methods collapse data when calculating the gradients.

In Chapter 6 we adapt our performance modeling method to the off-line quality control problem. Again, we approximate the circuit performances by functions of all the inputs to the circuit simulator, collect data from a statistically designed experiment, and fit the performance models. For a fixed set of designable parameters, the fitted models are used to predict the Taguchi loss statistic, rather than parametric yield. The loss statistic is numerically optimized with respect to the designable parameters. We give a circuit example in which the Taguchi objectives are met with about one-third of the runs. Finally, in Chapter 7 we provide some conclusions along with some suggestions for future research.

CHAPTER 2.

STATISTICAL MODELING OF THE MOS TRANSISTOR

A major goal of statistical circuit design is to predict, from a known distribution of the inputs to the circuit simulator, the distribution of the output circuit performances. In MOS integrated circuits these statistical inputs are the device parameters, such as threshold voltage and oxide capacitance. Typically, empirical distributions of the device parameters are obtained by test structure measurements and parameter extraction. Statistical circuit analysis by direct Monte Carlo circuit simulator runs is computationally too expensive, however. To analyze a circuit with a small number of runs, the distribution of the device parameters should be sampled effectively.

The traditional best- and worst-case analyses approach [10] to statistical circuit design estimates the "range" of the performances by circuit simulations at the "best-" and "worst-case" files. Best- and worst-case analyses do not estimate the performance distribution, however. In this chapter, we present a new statistical model to represent the distribution of the device parameters [11]. The model in [11] assumes that inter-die variations in the circuit performances are due to a small subset of critical parameters, and models the parameter correlations by fitted quasi-physical equations. This screening and modeling of MOSFET parameters [11] significantly reduce the problem dimension. Section 2.1 outlines the best- and worst-case analysis method. The statistical model of the device parameters is described in Section 2.2. A discussion in Section 2.3 concludes this chapter.

2.1. Best- and Worst-case Analyses

The best- and worst-case analyses approach is the traditional method for statistical circuit analysis. It estimates the "range" of the performances from a small number of circuit simulations (usually three to five runs). We denote the device parameters, such as threshold voltage and oxide capacitance, by $\mathbf{P} = (p_1, \dots, p_q)$. The commonly used methods to obtain the best- and worst-case files are

- (a) *One-at-a-time method*: Each p_i that leads to a better or worse performance is chosen independently. Typically, the chosen value is the mean of $p_i \pm 2$ or 3 standard deviation. This method gives very conservative best- and worst-case performances, because the probability of a MOSFET falling outside the rectangular region bounded by the best- and worst-case p_i 's is extremely small.
- (b) *Fast and slow MOSFETs*: Select "fast" and "slow" transistors from the test structures, and extract their parameters for circuit simulation. For digital circuits, the fast and slow MOSFETs are expected to give the best and worst performances. This method gives less conservative best- and worst-case estimates than the first approach because it incorporates the correlations between the p_i 's. The choice of the MOSFET samples remains a problem, however.
- (c) *Principal component approach* [12]: Transform the correlated parameters p_1, \dots, p_q into independent variables ξ_1, \dots, ξ_q by principal components, and then select the ξ_i , $i = 1, \dots, q$ for the circuit simulation. The problem of MOSFET parameter correlations is eliminated by the transformation.

(d) *Independent process parameters* [10]: Select the combinations of the manufacturing process parameters that lead to the best- and worst-case performances, and generate the MOSFET parameters from a process/device simulator, such as FABRICS [13]. Working on the independent process parameters avoids the analysis of the MOSFET parameters.

Best- and worst-case circuit simulations do not provide the circuit performance distributions. Moreover, there exist no formal methods to validate the choice of the best and worst-case files. Clearly, an improved representation of the device parameter distribution is needed.

2.2. Critical MOSFET Parameters

Yang and Chatterjee [11] observed that a small number of *critical MOSFET parameters* can account for most of the variability in the circuit behavior. Four critical MOSFET parameters were identified as [11]

- transistor channel width reduction (ΔW),
- transistor channel length reduction (ΔL),
- gate oxide thickness (t_{ox}), and
- flatband voltage (V_{fb}).

These critical parameters take independent Gaussian distributions. It has been shown [14] that currents and capacitances of MOS transistors are sensitive to the four critical parameters. Moreover, the sensitivities to the other device parameters are calculated to be at least an order of magnitude smaller [1].

The non-critical device parameters in P , such as

- drain and source resistances,
- body effect coefficient, and
- channel length modulation,

are approximated by quasi-physical equations of the critical parameters. We called the set of quasi-physical equations a *statistical MOSFET model*. From device physics and the charge sharing concept [15], it is recognized that most device parameters have an inverse dependency on the transistor channel widths and lengths. Thus, the model

$$p_i = p_{oi} + \frac{p_{wi}}{W - \Delta W} + \frac{p_{Li}}{L - \Delta L}, \quad (2.1)$$

where W and L are the transistor channel width and length, p_{oi} is the nominal value of p_i , and p_{wi} and p_{Li} represent the W and L dependencies. The exceptions to Eq. (2.1) are the transistor gain [1]:

$$\beta_m = \mu C_{ox} \frac{(W - \Delta W)}{(L - \Delta L)}, \quad (2.2)$$

where μ is the mobility and C_{ox} is the gate oxide capacitance; and the mobility degradation

$$\theta = \theta_o + \frac{\theta_w(W - \Delta W)}{(L - \Delta L)} + \frac{\theta_L}{(L - \Delta L)} + \frac{\theta_m}{(W - \Delta W)(L - \Delta L)}, \quad (2.3)$$

where θ_o is the mobility reduction due to the vertical electrical field in the channel and surface scattering, and θ_w , θ_L , and θ_m are the W and L dependency coefficients. Measured

data of the device parameters are fitted to the quasi-physical equations. The variability in a parameter that is not explained by the model will be ignored [11].

Liu and Singhal [16] suggested an alternative representation of the MOSFET parameters. They derived seven critical process parameters, $\mathbf{Z} = (z_1, \dots, z_7)$, from line-monitoring measurements

- offset in diffused line width (D_W),
- offset in poly width (D_L),
- oxide thickness (t_{ox}),
- flatband voltage (V_{fb}),
- substrate doping concentration (N_{SUB}),
- surface mobility (μ), and
- lateral junction depth (D_J).

The variation of each z_j is taken as a Gaussian distribution. The proposed statistical model is [16]

$$p_i = f_i(z_1, \dots, z_7) + g_i(L_{EFF}) + \sum_{j=1}^7 \left[a_{ij} z_j + \frac{b_{ij}}{z_j} \right] + \epsilon_i. \quad (2.4)$$

The model consists of four components:

- (a) A function $f_i(z_1, \dots, z_7)$ that relates a device parameter to its geometries and the manufacturing process.
- (b) A nonlinear function $g_i(L_{EFF})$ of the effective channel length L_{EFF} .
- (c) An expression $\sum_{j=1}^7 \left[a_{ij} z_j + \frac{b_{ij}}{z_j} \right]$, with unknown constants a_{ij} and b_{ij} , from multi-parameter linear regression [17].
- (d) A random error ϵ_i that models the statistical variation of p_i that is not explained by the first three components in Eq. (2.4). In the implementation, ϵ_i is generated by a Gaussian random number generator.

2.3. Discussion

The quasi-physical statistical MOSFET model in [1] is used in our yield prediction and optimization examples. Little justification has been introduced to support the more complex model in [16].

To restrict the number of independent statistical variables, the parameter variations within a die are neglected in [1] and [16]. Intra-die variations can be important in high performance analog circuits that require closely matched devices. The statistical analysis of circuits with parameter mismatches is considered in [18].

More investigation is needed to evaluate the validity of the statistical MOSFET models. The critical parameters in [1,16] are selected subjectively, and rigorous methods have not been used. A survey of screening methods and some examples are given in [19]. The step-wise regression and rank regression methods in [20,21] may be useful.

CHAPTER 3.

PARAMETRIC YIELD PREDICTION

Parametric yield prediction has become an increasingly important problem in the design of MOS integrated circuits. In the last decade, VLSI device feature sizes have decreased dramatically. However, the statistical variations of the device parameters have not been scaled down in proportion. As a result, the circuit performances become even more sensitive to statistical variations, which may lead to low parametric yield. To ensure acceptable yield, these statistical variations must be considered in the circuit design procedure.

A major bottleneck in parametric yield prediction is the high cost of the circuit simulation. The classical Monte Carlo method [22] is very expensive because it predicts the yield from a large number of circuit simulator runs. To reduce the simulation cost, a statistical modeling method is introduced in [1]. This approach approximates the circuit performances, such as gain and propagation delay time, by fitted models of four critical MOSFET parameters. These fitted models, estimated from five circuit simulations, act as computationally inexpensive surrogates of the circuit simulator in the Monte Carlo simulation. The study in [1] found many examples in which this performance modeling approach predicts yields fairly accurately. However, many important aspects in statistical modeling, such as model assessment and selection of the inputs to run the circuit simulator, have not been considered [1].

In this chapter we present an improved circuit performance modeling method for parametric yield prediction. As in [1], our method assumes that the circuit performances can be approximated by fitted models of a small subset of critical parameters, and perform Monte

Carlo simulation with these fitted models to predict yield. Unlike [1], the data in our method are collected according to a statistically designed experiment, which ensures all the relevant statistical information are gathered from a small number of circuit simulator runs. Furthermore, we assess the prediction capability of each fitted model before it is used to predict yield. Section 3.1 reviews the Monte Carlo and the statistical performance modeling methods. Section 3.2 summarizes the related research in statistics. Section 3.3 outlines our strategy. Section 3.4 illustrates our method with circuit examples. The discussion in Section 3.5 concludes this chapter.

3.1. Previous Parametric Yield Prediction Methods

The parametric yield of a MOS VLSI circuit depends on a large number of factors: the designable parameters, the operating conditions, the distribution of the device parameters, and the specified performance criteria. The designable parameters, such as the drawn transistor channel widths and lengths, are fixed in parametric yield prediction. (The case in which parametric yield is optimized with respect to the designable parameters is considered in Chapter 4.) The empirical distribution of the device parameters, such as channel threshold voltage and body effect coefficient, are usually obtained from test structure measurements.

We denote the circuit performances, such as gain or propagation delay time, by $\mathbf{Y} = (y_1, \dots, y_d)$. The device parameters are denoted by $\mathbf{P} = (p_1, \dots, p_q)$. To simplify notation we do not introduce separate symbols for the operating conditions, but they are used later in Chapter 4.

Given a set of constraints on the performances, let $I(y_1, \dots, y_d) = 1$ if all the constraints are met, and 0 otherwise. The parametric yield is the percentage of the manufactured circuits with $I = 1$:

$$\Phi = \int I[y_1(\mathbf{P}), \dots, y_d(\mathbf{P})] d\Theta(\mathbf{P}) * 100\%, \quad (3.1)$$

where Θ is the distribution of \mathbf{P} .

Existing methods for yield prediction usually can be classified into Monte Carlo methods [18,22,23] or statistical modeling methods [1,5].

3.1.1. Monte Carlo methods

Monte Carlo circuit simulation, or simple random sampling, is probably the most widely used method for parametric yield prediction [18,22]:

STEP 1: Generate a (large) number of samples of the device parameters \mathbf{P}_i , $i = 1, \dots, N_{MC}$ from Θ .

STEP 2: Simulate the circuit performance $\mathbf{Y}(\mathbf{P}_i)$, for $i = 1, \dots, N_{MC}$.

STEP 3: Calculate the predicted yield as the percentage of samples that are acceptable:

$$\hat{\Phi} = \frac{1}{N_{MC}} \sum_{i=1}^{N_{MC}} I(\mathbf{Y}(\mathbf{P}_i)) * 100\%. \quad (3.2)$$

The confidence limits on $\hat{\Phi}$ are estimated from [23]:

$$Prob(|\hat{\Phi} - \Phi| \leq \psi \sigma(\hat{\Phi})) \approx 2F(\psi) - 1, \quad (3.3)$$

where

$$\sigma(\hat{\Phi}) = \left[\frac{\Phi(1 - \Phi)}{N_{MC} - 1} \right]^{1/2} \quad (3.4)$$

is the variance of the yield estimate, and $F(\psi)$ is the cumulative standard Gaussian distribution. Note the accuracy of $\hat{\Phi}$ is inversely proportional to the square root of the sample size N_{MC} .

The advantages of simple Monte Carlo circuit simulation are

- (i) It states the confidence limits on the predicted yield, $\hat{\Phi}$.
- (ii) The sample size N_{MC} is independent of the number of input parameters. As a result, all the device parameters can be varied in the circuit simulation, at no extra cost.
- (iii) It makes no restrictive assumption on the distribution of the device parameters, $\Theta(\mathbf{P})$.
- (iv) It makes no restrictive assumption on the relations between the inputs and outputs of the circuit simulator.

Due to the high cost of circuit simulation, the Monte Carlo method is too expensive for larger circuits. Another disadvantage of Monte Carlo analysis is that it does not give a physical relationship between the device parameters and yield. Nonetheless, we will treat the yield estimated from Monte Carlo circuit simulation as the "actual" yield to assess the relative accuracy of the other approaches.

Hocevar et al. [23] studied *variance reduction* methods to reduce the sample size. The strategies considered are correlated sampling, importance sampling, control variates, and stratified sampling. No practical and general reduction technique has been found [23].

3.1.2. Statistical performance modeling method

Cox, Yang, Manhant-Shetti, and Chatterjee [1] proposed a statistical performance modeling approach. Although the number of device parameters is large, it has been shown in Section 2.2 that only a small subset is critical. The uncontrollable inter-chip variations in the critical parameters are assumed to be independent random variables. For convenience they are taken as Gaussian. Device parameter variations within a chip are assumed to be negligible.

We denote the four critical parameters in [1] by $U = (u_1, \dots, u_4)$. The method involves four steps [1]:

STEP 1: Select five points U_1, \dots, U_5 and simulate the circuits at these points. The device parameters P are treated as functions of the four critical parameters (see Section 2.2).

STEP 2: Assume models

$$y_k = \beta_{0k} + \sum_{i=1}^4 \beta_{ik} u_i + \text{error}, \quad k = 1, \dots, d, \quad (3.5)$$

where $\beta_{0k}, \dots, \beta_{4k}$ are the unknown constants. Fit the data to model (3.5) to obtain the approximation $\hat{y}_k(U)$.

STEP 3: The *yield body* Λ is a region defined by the fitted models, $U \in \Lambda$ if $\hat{Y}(U)$ satisfies the constraints. The parametric yield is computed by numerical integration

$$\hat{\Phi} = \int_{U \in \Lambda} d\Gamma(U) * 100\%, \quad (3.6)$$

where $\Gamma(P)$ is the distribution of P .

STEP 4: For a more accurate estimate of Φ , perform extra simulations at points close to the boundary of Λ , and repeat steps 2 and 3.

Visvanathan [5] also advocated the performance modeling approach for statistical circuit analysis, and implemented his method in the CENTER/ADVICE program. The inputs to the circuit simulator are denoted by ξ_1, \dots, ξ_m . These ξ_i 's are the operating conditions and the critical process parameters in [16]. The assumed models are [5]

$$y_k = \beta_{0k} + \sum_{i=1}^m \beta_{ik} \xi_i + \sum_{i=1}^m \beta_{ik} \xi_i^2, \quad k = 1, \dots, d, \quad (3.7)$$

where β_{0k} , β_{ik} , etc. are unknown constants. The models are fitted to data from $2m+1$ circuit simulator runs: one run at the nominal values of ξ_1, \dots, ξ_m , and $2m$ runs where each ξ_i is varied to its "extreme" values.

The circuit performance modeling method [1,5] is attractive because it requires only a small number of circuit simulator runs, and the fitted models estimate the sensitivity coefficients $\partial y_k / \partial u_i$. However, these aspects in statistical modeling are not considered in [1,5]:

- (a) The statistical design of experiments have not been used to select the inputs to run the circuit simulator (see Appendix A).
- (b) The models are fitted from the minimum number of runs, and no attempt is made to assess their goodness of fit (see Appendix B).

- (c) The effects of representing the device parameters as functions of four critical MOSFET parameters are not known.

3.2. Uncertainty Analysis

In statistics, the analysis of the responses (outputs) of a computer code when the inputs are subjected to statistical variations is called *uncertainty analysis* [19], or *sensitivity analysis* [20,21]. Parametric yield prediction can be considered as an uncertainty analysis problem. The circuit performance models in [1] are commonly known as response surface models [24]. The equivalent terminologies in VLSI design and statistics are listed in Table 3.1.

Iman, Helton, and Campbell [20,21] considered the sensitivity analysis of a computer code known as the "pathways-to-man" model. Like parametric yield prediction, their goal is to predict from a known distribution of the inputs, the distribution of the responses from the computer code. They suggested the following analysis procedure: (i) Identify a (small) subset of critical inputs by Latin hypercube sampling [25]. (ii) Construct response surface models of the outputs by the stepwise regression and rank regression methods [17]. (iii) Generate the distribution of the responses by Monte Carlo simulation with the fitted (response surface) models.

Downing, Gardner, and Hoffman [19] also examined the response surface approach for the uncertainty analysis of computer codes. Their analysis method is similar to the one in [20]: (i) The critical inputs are identified by a fractional factorial experiment [26] that varies all the input parameters. (ii) Each response of interest is approximated by a function of

Table 3.1 Equivalent terminology in yield prediction and the response surface method.

| | VLSI Design | Statistics |
|----|--|--|
| 1. | circuit simulation | computer experiment |
| 2. | circuit simulation plan | design of experiment |
| 3. | circuit performance | response, output variable |
| 4. | inputs to the circuit simulator | predictors, input variables, experimental factors |
| 5. | designable parameters | controllable factors |
| 6. | device parameters | noise factors, uncontrollable parameters, statistical variations |
| 7. | macro-model, statistical performance model | empirical model, regression model, response surface model |

the critical inputs, while the non-critical inputs to the simulation code are fixed at their nominal values. Typically the approximating function is a second-order polynomial. (iii) The approximating function is fitted to data collected according to a central composite design [26]. The goodness of fit of a function to data is assessed by the R^2 -statistic (see Appendix B). The study in [19] showed that the response surface approach predicts the distributions of the responses fairly accurately in certain examples. However, it is emphasized that response surfaces should be used with caution, because it is difficult to evaluate the effects of dropping the non-critical inputs.

3.3. Proposed Parametric Yield Prediction Method

The proposed method combines the device parameter modeling method [1] and the response surface method [24]. The statistical MOSFET models (device parameter equations) [1] allow us to reduce the dimension of the problem without compromising the accuracy of the yield prediction. We prefer the systematic response surface method over the empirical curve-fitting strategy in [1].

Denote the critical parameters by U and the circuit performances by y_k , $k = 1, \dots, d$. The broad strategy of our yield prediction algorithm involves four steps:

STEP 1: Assume model

$$y_k = f_k (U) + \text{error}, \quad k = 1, \dots, d, \quad (3.8)$$

for each circuit performance of interest, y_k .

STEP 2: Design an experiment and simulate the circuits at the experimental design points.

The non-critical device parameters are treated as functions of the critical inputs.

STEP 3: Fit the models (3.8) to obtain the approximation $\hat{y}_k(\mathbf{U})$, and improve the model if necessary.

STEP 4: The parametric yield is estimated by Monte Carlo simulation. First generate a (large) sample, \mathbf{U}_i , $i = 1, \dots, N_{MC}$, from the distribution of \mathbf{U} . Then compute

$$\hat{y}_k(\mathbf{U}_i), \quad i = 1, \dots, N_{MC}; k = 1, \dots, d. \quad (3.9)$$

The estimated parametric yield is the percentage of samples that are acceptable:

$$\hat{\Phi} = \frac{1}{N_{MC}} \sum_{i=1}^{N_{MC}} I(y_1(\mathbf{U}_i), \dots, y_d(\mathbf{U}_i)) * 100\%. \quad (3.10)$$

Detailed implementation of our experimental design and model assessment procedures is given in the circuit examples.

3.4. Yield Prediction Examples

Our method for statistical performance modeling and parametric yield prediction is illustrated through the following examples.

Example 1. NMOS chain of Inverters

We consider a chain of 20 NMOS inverters. The goals are to assess the adequacy of the statistical MOSFET model with four critical parameters and to fit an accurate prediction model of the performances. The modeled performances are

y_1 = the power dissipation [27],

y_2 = the output delay time.

The critical parameters are

u_1 = channel width reduction (ΔW),

u_2 = channel length reduction (ΔL),

u_3 = gate oxide capacitance (C_{ox}),

u_4 = channel flatband voltage (V_{fb}).

The factors u_1, \dots, u_4 are normalized to the range $[-1, +1]$, and their distribution functions are four independent Gaussian variables with mean zero, and $\sigma = 1/2$. We denote the four-dimensional experimental region $[-1, 1]^4$ by R .

The assumed model is

$$y_k = \beta_{0k} + \sum_{i=1}^4 \beta_{ik} u_i + \text{error}, \quad k = 1, 2, \quad (3.11)$$

where β_{0k}, β_{ik} , are the unknown regression coefficients.

Denote the design points by $U_i, i = 1, \dots, N_{OBS}$. At a design point U_i , the device parameter

$$p_{ij} = g_j(U_i) + e_{ij}, \quad j = 1, \dots, q, \quad (3.12)$$

where g_j is a systematic function that approximates the dependency of a parameter on U , and e_{ij} is a Gaussian random number that models the variation not determined by U . We

duced this random error into the experiment to assess the sufficiency of u_1, \dots, u_4 . In [1] there is no residual random variation, and $e_{ij} = 0$.

The number of runs in the experiment is fairly arbitrary: a minimum of 5 runs is needed, and 3 (replicated) runs are taken at the center of the design region R to assess potential lack of fit. The ACED package [28] was used to obtain the design (see Section A.1). This package can construct experiments according to various optimality criteria. The mean-squared error criterion in ACED, which addresses both the sampling error and bias in prediction arising from model inadequacy, is appropriate. The robust design criterion that chooses a compromise between the sampling and bias error is used. The ACED generated experimental design is listed in Table 3.2, along with the data.

Table 3.2 Experimental design and data of inverter chain (four critical parameters).

| Run | u_1 | u_2 | u_3 | u_4 | y_1 (mW) | y_2 (ns) |
|-----|-------|-------|-------|-------|------------|------------|
| 1 | -1 | -1 | -1 | +1 | 1.62 | 24.1 |
| 2 | -1 | +1 | +1 | +1 | 3.08 | 10.8 |
| 3 | -1 | +1 | +1 | -1 | 3.50 | 10.1 |
| 4 | 0 | 0 | 0 | 0 | 2.32 | 15.2 |
| 5 | 0 | 0 | 0 | 0 | 2.35 | 15.8 |
| 6 | 0 | 0 | 0 | 0 | 2.22 | 16.7 |
| 7 | +1 | -1 | -1 | -1 | 1.69 | 21.5 |
| 8 | +1 | -1 | +1 | +1 | 2.10 | 17.6 |
| 9 | +1 | +1 | -1 | +1 | 1.83 | 19.7 |

We fit the data to model (3.11) with the RSREG procedure in SAS [29] and assess the prediction capability of fitted model, \hat{y}_1 and \hat{y}_2 , by the F-test procedure in [30] (see Section B.1). Tables 3.3 (a) and (b) list the regression statistics, along with the fitted equations. The F-statistics estimate the ratios between the variations "explained" by a model and the error. They are used to test if

- (a) the "range" of values predicted by a fitted model is substantially larger than the standard error,
- (b) the lack of fit is insignificant.

First, condition (a) is checked at a significance level $\alpha = 0.05$, whether the variation "explained" by \hat{y}_1 is at least $\gamma_o^2 = 4$ times larger than its standard error. In this case, the F-statistic of the fitted model ($F_1 = 148.86$) is substantially larger than the critical value of the corresponding noncentral F-distribution ($F_{1,cr} = F'_{0.05;4,4,4} = 29.8$). This suggests the fitted linear equation of u_1, \dots, u_4 adequately accounts for the variations in y_1 .

Next, we check condition (b) for the adequacy of Eq. (3.11). The lack of fit F-statistic ($F_2 = 1.5$) is substantially smaller than the critical value of the corresponding F-distribution ($F_{2,cr} = F_{0.05;2,2} = 19.0$), suggesting the lack of fit is insignificant. We conclude the model is adequate (case 1 in Section B.1). Following the same procedure, we also conclude that the fitted delay equation is adequate.

The model predicted a maximum (worst-case) power of 3.5 mW in R , with a 95% confidence limit of ± 0.2 mW. At this point, the parametric yield can be estimated by Monte Carlo simulation with the fitted models. This procedure is described in the next example.

Table 3.3 (a) ANOVA table for the power of inverter chain (four critical parameters).

| Source | SS | df | MS | F-ratio | Test Statistic |
|-------------|--------|----|--------|---------------|-------------------|
| Model | 3.122 | 4 | 0.781 | $F_1 = 148.8$ | $F_{1,cr} = 29.8$ |
| Error | 0.021 | 4 | 0.005 | | |
| Lack of Fit | 0.0126 | 2 | 0.0065 | $F_2 = 1.125$ | $F_{2,cr} = 19.0$ |
| Pure Error | 0.0083 | 2 | 0.0042 | | |

$$\hat{y}_1 = 2.35 - 0.19u_1 + 0.29u_2 + 0.43u_3 - 0.21u_4$$

Table 3.3 (b) ANOVA table for the delay of inverter chain (four critical parameters).

| Source | SS | df | MS | F-ratio | Test Statistic |
|-------------|-------|----|-------|---------------|-------------------|
| Model | 163.5 | 4 | 40.88 | $F_1 = 32.3$ | $F_{1,cr} = 29.8$ |
| Error | 5.06 | 4 | 1.265 | | |
| Lack of Fit | 3.912 | 2 | 1.956 | $F_2 = 3.412$ | $F_{2,cr} = 19.0$ |
| Pure Error | 1.147 | 2 | 0.573 | | |

$$\hat{y}_2 = 16.6 + 0.3u_1 - 2.5u_2 - 3.6u_3 + 1.2u_4$$

Example 2. Critical Path of 32-Bit Domino CMOS ALU

The second example is the critical path of a 32-bit domino CMOS ALU circuit shown in Figure 3.1 [31]. The circuit performances of interest are:

y_1 = the power dissipation,

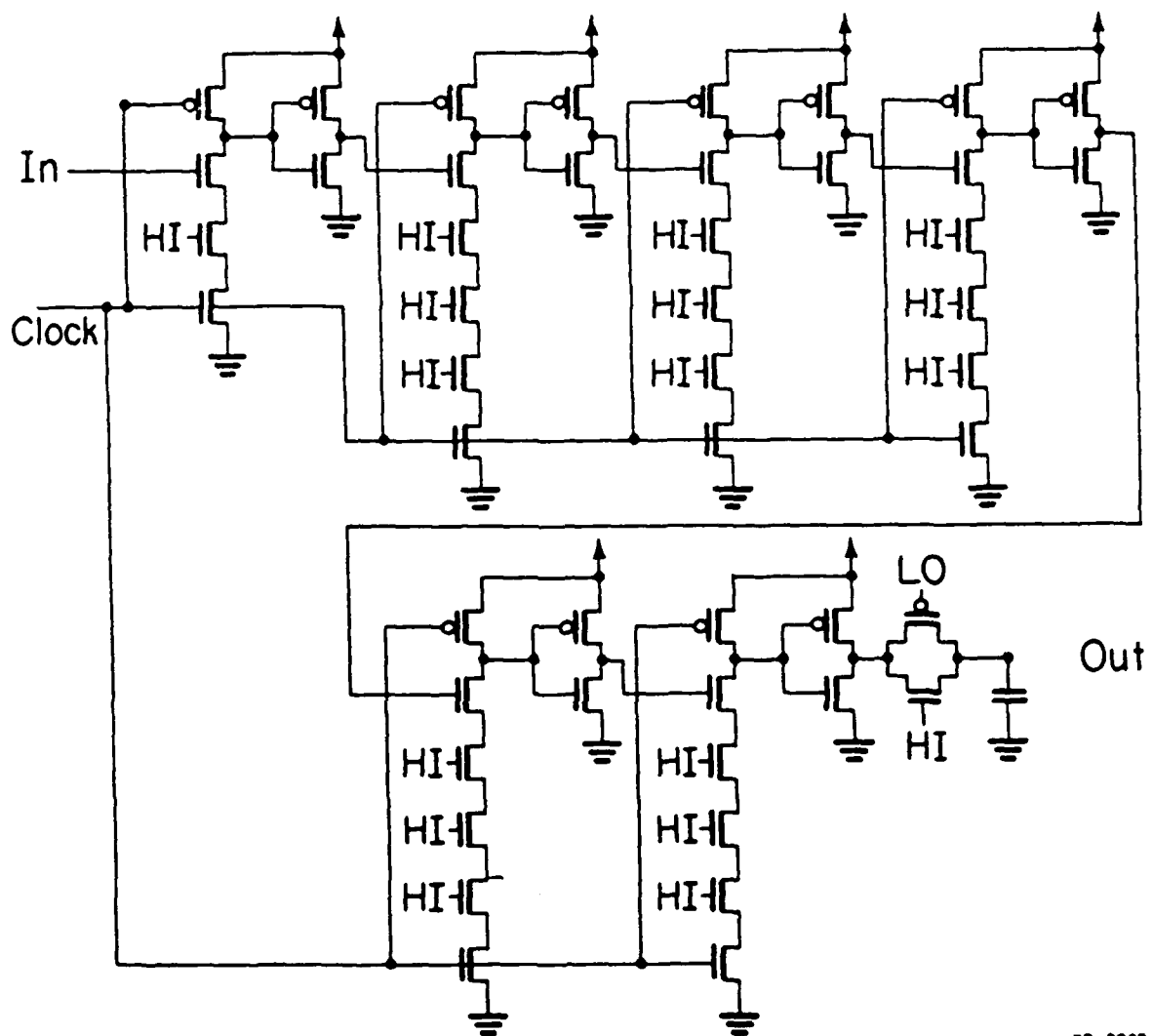
y_2 = the output delay time.

Table 3.4 shows the experimental design chosen by ACED, along with the data. The assumed model is identical to (3.11). Tables 3.5 (a) and (b) show the regression statistics, along with the fitted equations.

In the case of power dissipation, the F-test show that the range of values predicted by \hat{y}_1 is much larger than the error ($F_1 = 36.1 > F_{1,cr} = 29.8$), and the lack of fit is insignificant ($F_2 = 2.0 < F_{2,cr} = 19.0$). The test results suggest the model is adequate (case 1 in Section B.1).

In the case of output delay time, the F-tests show that the range of values predicted by \hat{y}_2 is not substantially larger than the standard error ($F_1 = 20.9 < F_{1,cr} = 29.8$), and the lack of fit is insignificant ($F_2 = 1.2 < F_{2,cr} = 19.0$). We conclude

- The fitted linear equation of u_1, u_2, u_3, u_4 , is not an adequate predictor of the delay time.
- The linear equation should approximate the actual delay adequately.
- The error due to uncertainties in the fitted model is important.



FP-9062

Figure 3.1 Critical path through a 32-bit Domino CMOS ALU.

Table 3.4 Experimental design and data of ALU (four critical parameters).

| Run | u_1 | u_2 | u_3 | u_4 | y_1 (mW) | y_2 (ns) |
|-----|-------|-------|-------|-------|------------|------------|
| 1 | -1 | -1 | -1 | +1 | 6.12 | 80.7 |
| 2 | -1 | +1 | +1 | +1 | 7.32 | 51.0 |
| 3 | -1 | +1 | +1 | -1 | 5.67 | 40.8 |
| 4 | 0 | 0 | 0 | 0 | 6.04 | 62.5 |
| 5 | 0 | 0 | 0 | 0 | 6.20 | 65.1 |
| 6 | 0 | 0 | 0 | 0 | 5.75 | 55.6 |
| 7 | +1 | -1 | -1 | -1 | 5.12 | 80.8 |
| 8 | +1 | -1 | +1 | +1 | 9.16 | 96.2 |
| 9 | +1 | +1 | -1 | +1 | 6.25 | 61.2 |

Therefore, we conclude that either the number of observations N_{OBS} is too small or that the set of critical parameters used is incomplete (case 2 in Section B.1).

It is observed that the domino CMOS circuit uses a long chain of NMOS transistors connected in series; the effect of the back-gate bias on the output delay time should be significant. However, the substrate doping N_{SUB} , which strongly affects the body-effect coefficient, is not included in the four-parameter model (3.11). Therefore, we include a fifth critical MOSFET parameter

u_5 = substrate doping (N_{SUB}) .

The assumed model is

$$y_2 = \beta_{0k} + \sum_{i=1}^5 \beta_{ik} u_i + error, \quad k = 1, 2, \quad (3.13)$$

Table 3.5 (a) ANOVA table for the power of ALU (four critical parameters).

| Source | SS | df | MS | F-ratio | Test Statistic |
|-------------|-------|----|-------|--------------|-------------------|
| Model | 11.02 | 4 | 2.76 | $F_1 = 36.1$ | $F_{1,cr} = 29.8$ |
| Error | 0.305 | 4 | 0.08 | | |
| Lack of Fit | 0.202 | 2 | 0.101 | $F_2 = 2.0$ | $F_{2,cr} = 19.0$ |
| Pure Error | 0.103 | 2 | 0.051 | | |

$$\hat{y}_1 = 6.19 + 0.46u_1 - 0.39u_2 + 1.06u_3 + 0.95u_4$$

Table 3.5 (b) ANOVA table for the delay of ALU (four critical parameters).

| Source | SS | df | MS | F-ratio | Test Statistic |
|-------------|--------|----|-------|--------------|-------------------|
| Model | 2243.0 | 4 | 560.8 | $F_1 = 20.9$ | $F_{1,cr} = 29.8$ |
| Error | 107.2 | 4 | 26.8 | | |
| Lack of Fit | 58.5 | 2 | 29.3 | $F_2 = 1.2$ | $F_{2,cr} = 19.0$ |
| Pure Error | 48.7 | 2 | 24.4 | | |

$$\hat{y}_2 = 64.5 + 6.2u_1 - 15.9u_2 + 1.6u_3 + 6.4u_4$$

Table 3.6 Experimental design and data of ALU (five critical parameters).

| Run | u_1 | u_2 | u_3 | u_4 | u_5 | y_1 (mW) | y_2 (ns) |
|-----|-------|-------|-------|-------|-------|------------|------------|
| 1 | -1 | -1 | +1 | -1 | -1 | 5.76 | 61.0 |
| 2 | -1 | -1 | +1 | +1 | +1 | 7.42 | 71.4 |
| 3 | -1 | +1 | -1 | -1 | -1 | 4.76 | 46.9 |
| 4 | -1 | +1 | -1 | +1 | +1 | 6.30 | 57.2 |
| 5 | 0 | 0 | 0 | 0 | 0 | 6.40 | 68.4 |
| 6 | 0 | 0 | 0 | 0 | 0 | 5.78 | 59.4 |
| 7 | 0 | 0 | 0 | 0 | 0 | 6.29 | 65.1 |
| 8 | +1 | -1 | -1 | -1 | +1 | 5.27 | 90.6 |
| 9 | +1 | -1 | -1 | +1 | -1 | 6.30 | 87.5 |
| 10 | +1 | +1 | +1 | -1 | +1 | 5.97 | 52.0 |
| 11 | +1 | +1 | +1 | +1 | -1 | 7.10 | 50.6 |

An experimental design for u_1, \dots, u_5 with $N_{OBS} = 11$ runs is generated by ACED. Table 3.6 shows the experimental design, along with the data. The fitted models and the regression statistics are shown in Table 3.7.

In this case $F_1 = 46.7$ is larger than $F_{1,cr}' = F'_{0.05;5,5,4} = 23.2$, and $F_2 = 0.02$ is smaller than $F_{2,cr} = F_{0.05;3,3} = 19.0$ for the delay, which indicates that the delay model is now adequate with five parameters. It should be noted that the statistical variation of N_{SUB} is technology dependent and, for some technology, the variation of N_{SUB} may be negligible.

The distribution of the output delay time, obtained from 10,000 Monte Carlo runs of the fitted model, is shown in Figure 3.2. The number $N_{MC} = 10,000$ is chosen simply to ensure a sufficient number of samples. The computer time to evaluate 10,000 sample points with the fitted models is very small compared with the time for the circuit simulation. An empirical distribution of the delay, obtained from 300 runs of the circuit simulator with all the input

Table 3.7 (a) ANOVA table for the power of ALU (five critical parameters)

| Source | SS | df | MS | F-ratio | Test Statistic |
|-------------|-------|----|-------|--------------|-------------------|
| Model | 5.44 | 5 | 1.09 | $F_1 = 23.8$ | $F_{1,cr} = 23.2$ |
| Error | 0.23 | 5 | 0.05 | | |
| Lack of Fit | 0.012 | 3 | 0.004 | $F_2 = 0.04$ | $F_{2,cr} = 19.2$ |
| Pure Error | 0.217 | 2 | 0.109 | | |

$$y_2 = 6.12 + 0.05u_1 - 0.08u_2 + 0.45u_3 + 0.67u_4 + 0.13u_5$$

Table 3.7 (b) ANOVA table for the power of ALU (five critical parameters).

| Source | SS | df | MS | F-ratio | Test Statistic |
|-------------|--------|----|-------|--------------|-------------------|
| Model | 1978.2 | 5 | 395.6 | $F_1 = 46.7$ | $F_{1,cr} = 23.2$ |
| Error | 42.4 | 5 | 8.9 | | |
| Lack of Fit | 1.1 | 3 | 0.4 | $F_2 = 0.02$ | $F_{2,cr} = 19.2$ |
| Pure Error | 41.3 | 2 | 20.7 | | |

$$\hat{y}_2 = 64.5 + 5.5u_1 - 13.0u_2 - 5.9u_3 + 2.0u_4 + 3.2u_5$$

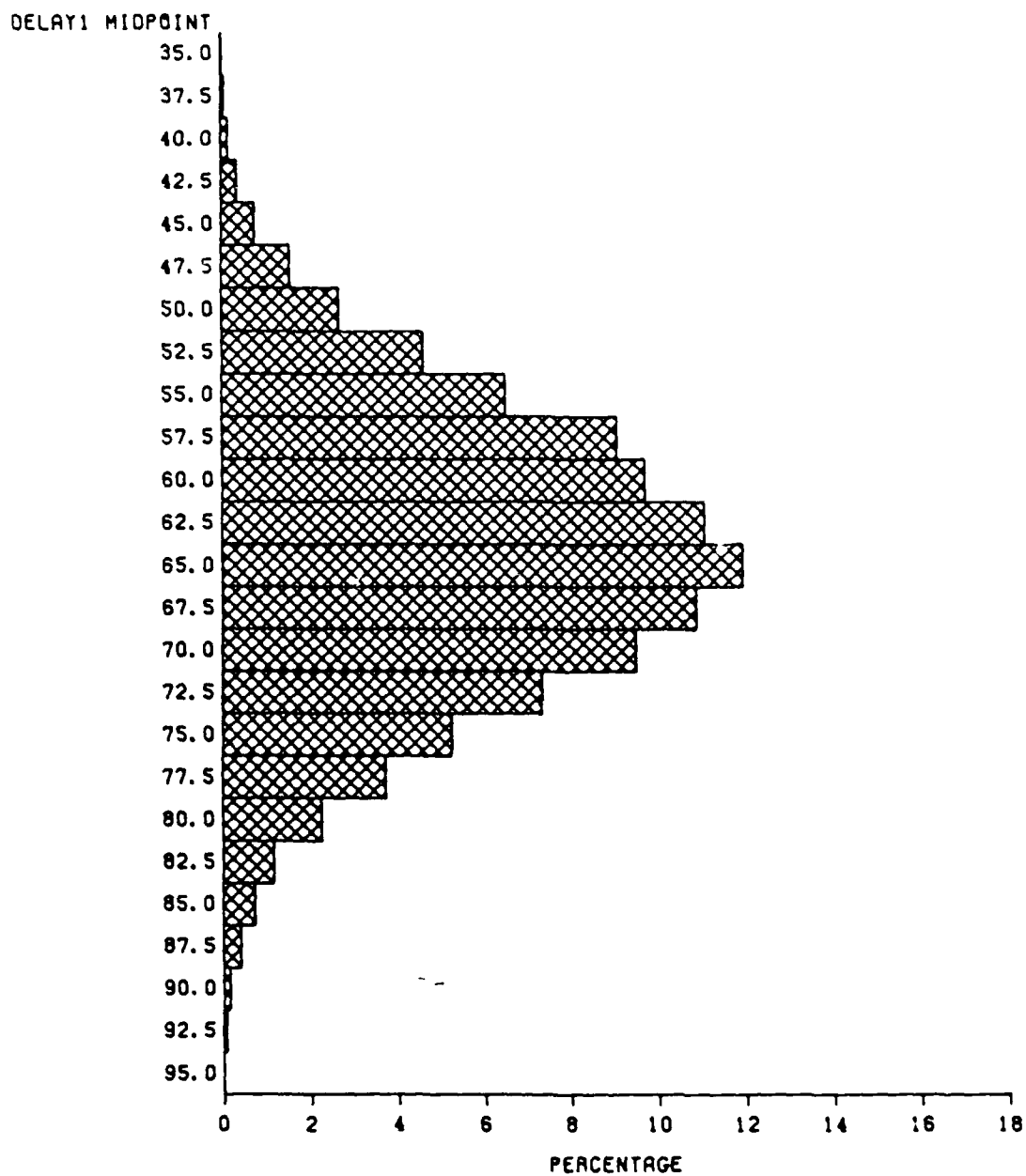


Figure 3.2 ALU delay distribution from 10,000 runs of the fitted model.

MOSFET parameters varying, is shown in Figure 3.3. The moments and percentage points of the two distributions are listed in Table 3.8. We use the distribution from the 300 circuit simulator runs as the "true" delay distribution because it is the best approximation that is available.

The fitted model predicts the distribution of the delay accurately. The differences between the means and standard deviations of the delay distributions are small (5%). However, the distribution from 10,000 runs with the model is symmetric and normal whereas the distribution from 300 circuit simulator runs has a skewness of 0.8. These discrepancies can be explained in part by the higher-order effects not modeled in (3.13) and declared insignificant by the lack of fit test.

The differences between the 75, 90, 95 and 99 percentile points of the delay distributions are consistently less than 10%, and there is a maximum error of +5.9% at the 99 percentile.

The parametric yields for various delay constraints are estimated from the performance distributions. The yield difference between the regression model and circuit simulator based simulation for delay constraints of 65 ns, 70 ns and 75 ns is consistently less than 12%.

Example 3. CMOS 4-Bit Full Adder

In this example we consider a CMOS 4-bit full adder consisting of 112 transistors. The modeled performance is

y = the output delay time of the most significant bit.

The five critical parameters u_1, \dots, u_5 are identical to those in Example 2. ACED is used to

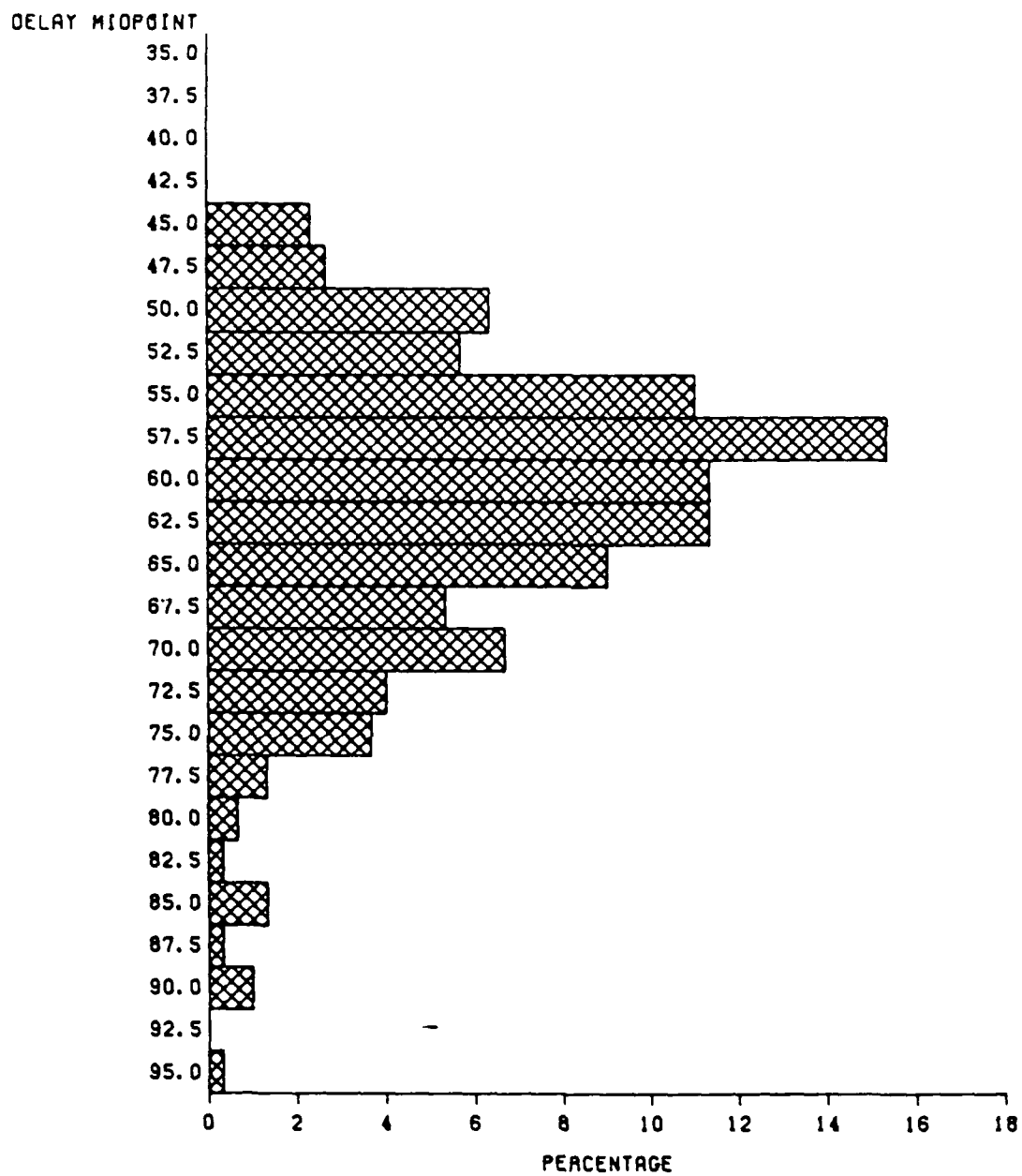


Figure 3.3 ALU delay distribution from 300 circuit simulator runs.

Table 3.8 Delay distribution of ALU.

| | Monte Carlo with Circuit Simulator | Monte Carlo with Fitted Model |
|----------------------------|---------------------------------------|----------------------------------|
| No. samples | 300 | 10000 |
| Mean | 61.4 | 64.5 |
| Std. Dev. | 8.96 | 8.48 |
| Variance | 80.2 | 71.9 |
| Skewness | 0.81 | 0.0 |
| Kurtosis | 1.05 | 0.0 |
| Median | 60.3 | 64.5 |
| 75 percentile | 66.3 | 70.2 |
| 90 percentile | 73.1 | 75.4 |
| 95 percentile | 77.6 | 78.5 |
| 99 percentile | 90.0 | 84.8 |
| $\hat{\Phi}_{65} \ddagger$ | 75% | 64% |
| $\hat{\Phi}_{70}$ | 87% | 83% |
| $\hat{\Phi}_{75}$ | 95% | 94% |

$\ddagger \hat{\Phi}_{T_{\max}}$ is the predicted parameteric yield when the delay constraint is T_{\max} (ns).

design a 16-run experiment of 5 factors. The delay distributions generated by 10,000 runs with the fitted model and 300 runs of the circuit simulator varying all the device parameters are compared in Table 3.9. It can be observed that the errors in the worst-case delay estimation are consistently less than 5% and the differences in yield estimates are less than 8%.

3.5. Concluding Remarks

In this chapter, an improved circuit performance modeling method for parametric yield prediction is presented. It is shown that our method predicts yield accurately with a relatively small number of circuit simulator runs.

We introduced the response surface method for the building of the circuit performance models. Our method, based on the design and analysis of experiments, allows accurate models of the circuit performances to be fitted from a small number of circuit simulator runs. Experimental design also allows the assessment of the predictive capabilities of the fitted models by a statistical F-test procedure.

A systematic procedure is introduced to verify whether the four critical parameters indeed cause most of the variations in the circuit performances. Two statistical F-tests are used to compare the variations due to the critical parameters with the variations due to other sources. As demonstrated through the 32-bit ALU circuit example, the four critical parameters $\{\Delta W, \Delta L, C_{ox}, V_{fb}\}$ may not be always sufficient. In this example engineering insight has been used to identify the fifth critical parameter N_{SUB} . This empirical method of parameter screening may be avoided if the systematic screening methods were used (see Section 2.3).

Table 3.9 Delay distribution of four-bit full adder.

| | Monte Carlo with Circuit Simulator | Monte Carlo with Fitted Model |
|--------------------|---------------------------------------|----------------------------------|
| No. samples | 300 | 10000 |
| Mean | 78.3 | 81.3 |
| Std. Dev. | 13.3 | 14.9 |
| Variance | 176.3 | 220.6 |
| Skewness | 0.45 | 0.0 |
| Kurtosis | 0.0 | 0.0 |
| Median | 78.1 | 81.4 |
| 75 percentile | 86.8 | 91.3 |
| 90 percentile | 97.6 | 100.3 |
| 95 percentile | 101.9 | 105.5 |
| 99 percentile | 114.6 | 115.2 |
| $\hat{\Phi}_{80}$ | 55% | 47% |
| $\hat{\Phi}_{90}$ | 81% | 73% |
| $\hat{\Phi}_{100}$ | 93% | 90% |

‡ $\hat{\Phi}_{T_{\max}}$ is the predicted parameteric yield when the
the delay constraint is T_{\max} (ns).

Based on our experience with these and other examples, a linear equation often adequately approximates the performance of digital circuits. With a sufficient set of critical parameters and careful modeling, the parametric yield and the distribution of the performances can be predicted accurately.

In the rest of the thesis we will assume the critical parameters to be sufficient. The other device parameters will be treated as deterministic functions of the critical parameters only. This eliminates the sampling errors from the fitted models. The design of computer-simulation experiments that has no random error is considered in Section A.2.

CHAPTER 4.

PARAMETRIC YIELD OPTIMIZATION

In this chapter we extend the circuit performance modeling method to parametric yield optimization. Like parametric prediction, the bottleneck in yield optimization is the cost of the circuit simulation. Over the past years many approaches have been proposed for yield optimization, typically many circuit simulator runs are required, and their practical applications are limited.

Yield gradient methods for parametric yield maximization [6, 32, 33] compute gradients with respect to the designable circuit parameters and then use steepest ascent to optimize the yield. Gradients only represent the local behavior of the yield function, however, and a possibly poor local maximum may be found (unless the yield function is simple). Yield gradient methods typically require a large number of circuit simulations. Data from many runs of the circuit simulator are required to estimate a single yield gradient, and several gradient computations are needed in the optimization.

In our approach, we model each circuit performance as a function of all parameters of interest: the designable parameters, the statistical variations, and the operating conditions. Data generated according to a statistical experiment are used to identify and fit these models. For fixed values of the designable parameters, the Monte Carlo yield is estimated with the approximating models acting as computationally cheap surrogates for the circuit simulator, as in Chapter 3. This estimated yield is numerically optimized. Section 4.1 outlines the steps in

our strategy, and Section 4.3 illustrates the detailed implementation in two CMOS analog circuit examples. Finally, Section 4.4 presents some concluding remarks.

4.1. Proposed Parametric Yield Maximization Method

The parametric yield of a MOS VLSI circuit depends on a large number of factors: the designable parameters, the operating conditions, the distribution of the uncontrollable statistical variations, and the specified performance criteria. The designable circuit parameters, such as the drawn transistor channel lengths and the aspect ratios of the P and N-channel transistors, can be specified by the circuit designers. We model the distribution of the device parameters by a small subset of critical parameters, as in Chapter 3. The device parameters are treated as functions of the critical parameters only, and there is no randomness in the experiment.

We denote the circuit performances of interest, such as gain or propagation delay time, by $Y = (y_1, \dots, y_d)$. Let $X = (x_1, \dots, x_p)$ denote the varying input parameters to the circuit simulator, all other inputs remaining fixed. In circuit simulation, each x_i can be manipulated to represent controllable adjustment of a designable circuit parameter and/or uncontrollable statistical variation. We write $x_i = c_i + u_i$ to differentiate between the controllable and uncontrollable portions. If an input parameter has no designable adjustment, then c_i has a fixed value and is ignored. Similarly, if there is no uncontrollable variation, then $u_i = 0$. Each y_k is, therefore, a function of $X = C + U$, where $C = (c_1, \dots, c_p)$ and $U = (u_1, \dots, u_p)$.

Given a set of constraints on the performances, let $I(y_1, \dots, y_d) = 1$ if all the constraints are met, and 0 otherwise. For given controllable parameters C , the parametric yield is the percentage of the manufactured circuits with $I = 1$:

$$\Phi(C) = \int I[y_1(C+U), \dots, y_d(C+U)] d\Gamma(U) * 100\%, \quad (4.1)$$

where Γ is the distribution of U .

The broad strategy for yield maximization involves six steps. The examples will illustrate details of their implementation.

STEP 1: Assume models for the performances y_k as functions of the circuit simulator inputs $X = C + U$:

$$y_k = f_k(X) + \text{error}. \quad (4.2)$$

The circuit simulator is deterministic, and the error term in (4.2) represents systematic departure of the assumed model $f_k(X)$ from the actual performance.

STEP 2: Design an experiment and simulate the circuits at the experimental design points.

STEP 3: Fit the models (4.2) to obtain the approximation $\hat{y}_k(X)$. Check if the model fits well, and improve the model if necessary. For example, a circuit performance that is too complex for accurate modeling can sometimes be identified as a composite function of several subcircuit performances; this is exploited in Section 4.2.

STEP 4: The parametric yield of a circuit design C , defined in Eq. (4.1), is estimated by Monte Carlo simulation. First, generate a (large) sample, U_i , $i = 1, \dots, N_{MC}$, from

the distribution Γ of the uncontrollable variations. Then, compute

$$\hat{y}_k(C + U_i), \quad i = 1, \dots, N_{MC}; \quad k = 1, \dots, d. \quad (4.3)$$

The estimated parametric yield is the percentage of samples that are acceptable:

$$\hat{\Phi}(C) = \frac{1}{N_{MC}} \sum_{i=1}^{N_{MC}} I[\hat{y}_1(C + U_i), \dots, \hat{y}_d(C + U_i)] * 100\%, \quad (4.4)$$

where I is defined in Eq. (4.1).

STEP 5: Maximize the estimated yield $\hat{\Phi}(C)$ with respect to C , and denote the resulting circuit design by C^* . We use the same Monte Carlo sample at each iteration of the optimization algorithm. Because $\hat{\Phi}(C)$ is not necessarily a smooth function of C , the simplex optimization algorithm [34] is used for optimization.

STEP 6: Obtain a "confirmatory" estimate of the yield at C^* . We use the method described in [3] which predicts the yield at C^* from a new, smaller experiment. The use of the circuit simulator for a Monte Carlo estimate of the yield at C^* would require an impractical number of runs.

In the examples below, at steps 1 and 3 we use quadratic regression models fit to the data by least squares. At step 2 the inputs are selected by a Latin hypercube - such designs were suggested by [25] for computer-simulation experiments. These simple tools lead to adequate accuracy here.

Polynomial models may not fit so well in other examples with more-complex performance functions. Also, the errors in model (4.2) are systematic and not the random errors

usually assumed when fitting by least squares [17]. Alternative implementations for the modeling, experimental design, and estimation steps are described in [35].

4.2. Application to Yield Maximization of CMOS Analog Circuits

Example A. CMOS comparator I

A CMOS two-stage comparator circuit is shown in Figure 4.1 [36]. The performances of interest and their initial constraints are

$$y_1 = \text{dc gain } (A_v > 1000),$$

$$y_2 = \text{propagation delay time } (T_1 < 170 \text{ ns}),$$

$$y_3 = \text{propagation delay time } (T_2 < 170 \text{ ns}).$$

The transient analysis of a comparator circuit is shown in Figure 4.2. The input voltages are denoted by V_+ and V_- , and the output is denoted by V_{out} . The time intervals T_1 and T_2 are the differences between the switching times of the input V_+ and the output voltage V_{out} .

The parameters used to represent the process variations and their $[-3\sigma, +3\sigma]$ ranges are

$$u_1 = \text{PMOS channel length reduction } (0.1 \mu m \leq \Delta L_P \leq 0.7 \mu m),$$

$$u_2 = \text{PMOS flatband voltage } (0.5 \text{ V} \leq V_{fbP} \leq 0.8 \text{ V}),$$

$$u_3 = \text{NMOS channel length reduction } (0.1 \mu m \leq \Delta L_N \leq 0.7 \mu m),$$

$$u_4 = \text{NMOS flatband voltage } (-0.6 \text{ V} \leq V_{fbN} \leq -0.4 \text{ V}),$$

$$u_5 = \text{gate oxide thickness } (37 \text{ nm} \leq t_{ox} \leq 43 \text{ nm}),$$

$$u_6 = \text{bias current variation } (-3 \mu A \leq \Delta I_B \leq +3 \mu A).$$

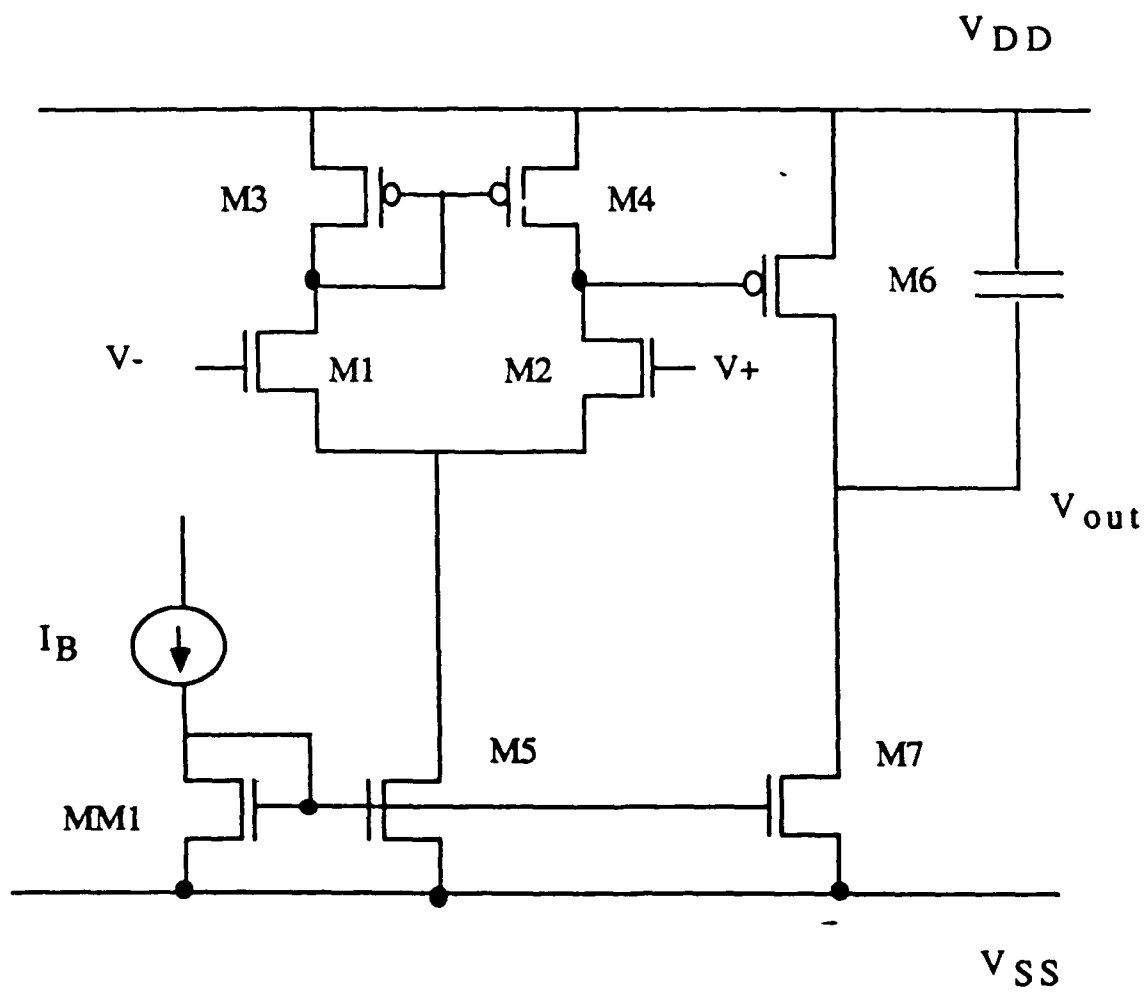


Figure 4.1 Two-stage CMOS comparator for Example A.

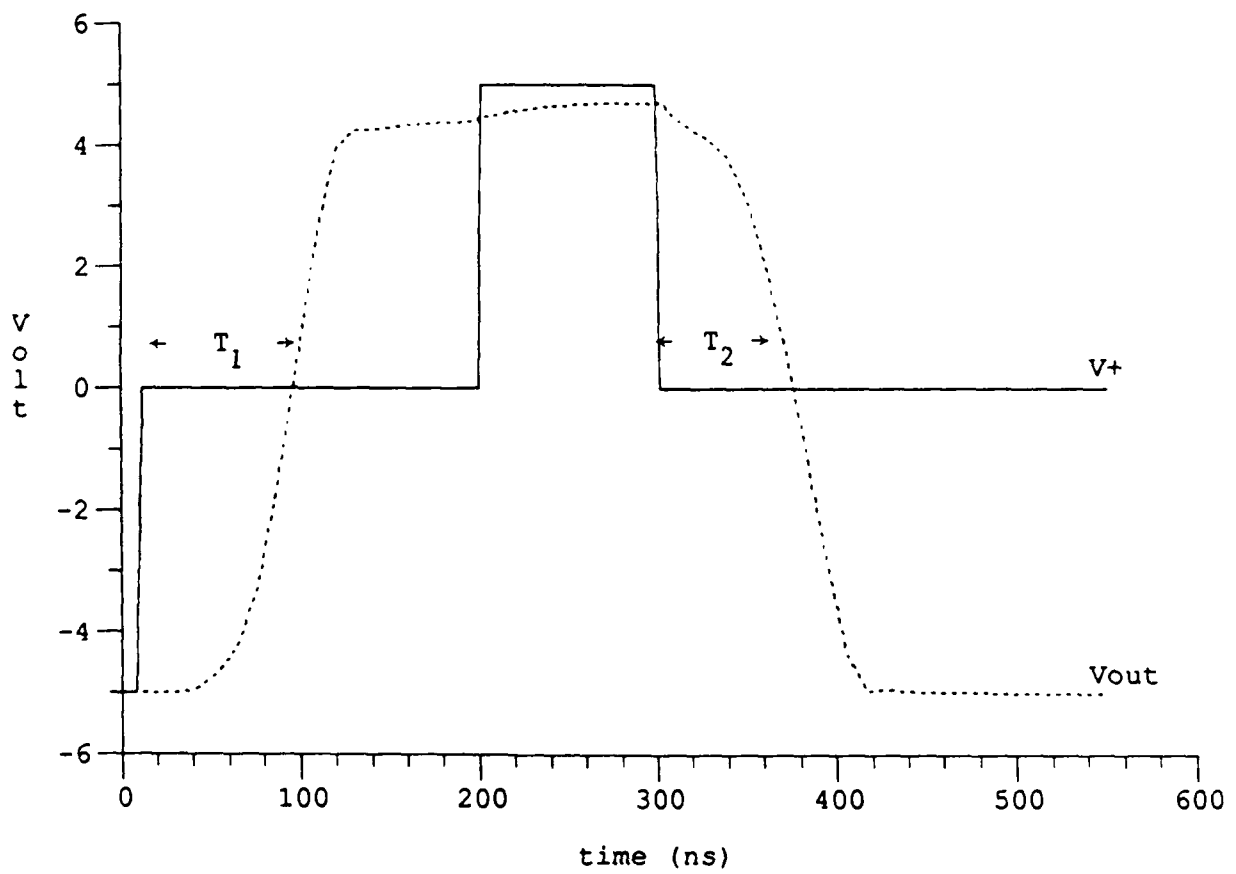


Figure 4.2 Transient response of CMOS comparator.

In addition there is a single, yet very demanding, operating condition:

v = input dc voltage, V_{DC} ($V_{DC} = -3.5, 0, \text{ or } 3.5 \text{ V}$).

For a circuit to be satisfactory, it must satisfy the constraints for (y_1, y_2, y_3) at all three levels of v .

To maximize the yield, we identify four designable parameters and determine their ranges empirically:

c_6 = nominal bias current ($40 \mu\text{A} \leq I_B \leq 60 \mu\text{A}$),

c_7 = width of transistors M1 and M2 ($150 \mu\text{m} \leq W(\text{M1, M2}) \leq 300 \mu\text{m}$),

c_8 = width of transistor M7 ($30 \mu\text{m} \leq W(\text{M7}) \leq 60 \mu\text{m}$),

c_9 = ratio of the widths of M7 and M6 ($1.6 \leq W(\text{M6})/W(\text{M7}) \leq 2.2$).

The parameter ranges are normalized as follows: $u_1, \dots, u_5, c_7, c_8$, and c_9 to $[-1, +1]$, and v to the discrete values $\{-1, 0, 1\}$. The bias current is made up of the nominal value, c_6 , and the variability, u_6 ; the total range of $I_B = c_6 + u_6$ is $37 \mu\text{A} \leq I_B \leq 63 \mu\text{A}$, which is then normalized to $[-1.3, +1.3]$. The distribution of U is that of six independent Gaussian variables with mean zero and standard deviation $\sigma = 1/3$.

Denote $(u_1, \dots, u_5, c_6 + u_6, c_7, c_8, c_9, v)$ by (x_1, \dots, x_{10}) . The assumed models are

$$y_k = \beta_{0k} + \sum_{i=1}^{10} \beta_{ik} x_i + \sum_{i=1}^{10} \sum_{j=i}^{10} \beta_{ijk} x_i x_j + \text{error}, \quad k = 1, 2, 3. \quad (4.5)$$

The constants β_{0k} , β_{ik} , and β_{ijk} , are (unknown) regression coefficients.

A 100-observation Latin hypercube experiment [25] is generated for the circuit simulation: a minimum of 66 runs is required to fit the model (4.5), and the remaining 34 degrees

of freedom permit some assessment of fit. Latin hypercube designs were developed for computer-simulation experiments and are easy to construct. The experimental design is generated as follows. Each x_i , except for x_4 and x_{10} , takes the equally spaced values $-99/100, -97/100, \dots, 99/100$, for the 100 runs, but in different random orders. Similarly, x_6 is generated in $[-1.3, +1.3]$ and x_{10} in the set $\{-1, 0, 1\}$. Thus, each variable is exercised over its range, and the variables are matched at random. The circuit descriptions are generated from the experimental design by iEDISON [37], and SPICE is used for circuit simulation. Table 4.1 lists part of the experimental design and the data.

For each of the three performances, the unknown coefficients in the model (4.5) are estimated by least squares based on the 100 observed runs. Since there is no random error in the circuit simulation, statistical testing of the fitted model is inappropriate. Nonetheless, we estimate the goodness of fit by R^2 [17] and R^2_{PRS} . The R^2 statistic measures the proportion of the variability in the data "explained" by the regression, while R^2_{PRS} is a modification that is useful for detecting possible model defects (see Appendix B).

The quadratic regression model for gain gives a bad fit with $R^2 = 0.914$, but $R^2_{PRS} = -0.02$. This clear indication of lack of fit is caused by 11 points with unusually low gains. For instance, at run number 96 the gain $A_v = 24 \ll 1000$. We examined the circuit simulation outputs at these anomalous observations, and found that either transistor M1 (2 cases) or M7 (9 cases) is not operating in the saturation region. The small-signal gain

$$A_v = \left[\frac{-g_{m1}}{g_{ds1} + g_{ds3}} \right] \left[\frac{-g_{m1}}{g_{ds6} + g_{ds7}} \right] \quad (4.6)$$

Table 4.1 Part of the 100-run Latin hypercube design for x_1, \dots, x_{10} and the observed y_1, \dots, y_5 for Example A.

| Run | x_1 | x_2 | x_3 | x_4 | x_5 | x_6 | x_7 | x_8 | x_9 | x_{10} | y_1 | y_2 | y_3 | y_4 | y_5 |
|-----|-------|-------|-------|-------|-------|-------|-------|-------|-------|----------|-------|-------|-------|-------|-------|
| 1 | 0.21 | 0.91 | -0.81 | 0.23 | -0.61 | 0.90 | 0.29 | 0.35 | -0.23 | 1 | 1649 | 67 | 52 | 0.4 | 5.9 |
| 2 | 0.95 | 0.67 | -0.07 | -0.61 | 0.05 | 1.00 | -0.01 | 0.23 | 0.41 | 1 | 1427 | 45 | 51 | 0.2 | 6.4 |
| 3 | -0.05 | -0.81 | 0.83 | 0.21 | -0.75 | -0.53 | 0.65 | -0.79 | -0.83 | -1 | 3469 | 109 | 129 | 7.4 | 0.2 |
| 4 | -0.17 | 0.31 | 0.35 | -0.87 | -0.93 | -0.19 | 0.93 | -0.87 | -0.71 | -1 | 3342 | 96 | 117 | 7.3 | 1.4 |
| 5 | -0.69 | 0.47 | -0.33 | -0.05 | -0.95 | 0.25 | -0.93 | 0.49 | 0.77 | -1 | 3063 | 112 | 220 | 7.5 | 1.2 |
| | | | | | | | | : | | | | | | | |
| | | | | | | | | : | | | | | | | |
| 96 | 0.63 | -0.25 | -0.23 | -0.77 | 0.77 | 1.18 | 0.17 | 0.55 | -0.09 | 1 | 24 | 37 | * | -0.0 | 5.6 |
| 97 | -0.19 | 0.87 | -0.19 | 0.15 | 0.19 | 1.26 | 0.83 | -0.45 | -0.93 | 1 | 1477 | 66 | 45 | 0.3 | 4.3 |
| 98 | 0.09 | -0.63 | 0.91 | -0.35 | 0.07 | -1.29 | -0.95 | 0.89 | -0.99 | -1 | 3044 | 120 | 209 | 7.4 | 0.8 |
| 99 | -0.73 | -0.65 | -0.49 | -0.07 | 0.41 | 0.66 | 0.23 | -0.39 | -0.01 | 0 | 1901 | 88 | 65 | 3.6 | 3.6 |
| 100 | 0.13 | 0.77 | 0.69 | 0.53 | 0.15 | -1.23 | -0.55 | 0.01 | -0.33 | -1 | 3543 | 113 | 203 | 7.8 | 0.8 |

Note: * indicates no response from the circuit

is low at these points, due to the large drain-to-source conductances of the unsaturated transistors M1 or M7 (g_{ds1} or g_{ds7}).

To cope with these problems, we construct two new models for the terminal voltages of M1 and M7:

$$y_4 = V_{ds1} - (V_{gs1} - V_{t1}) \quad (y_4 > 0 \text{ for M1 to be saturated}),$$

$$y_5 = V_{ds7} - (V_{gs7} - V_{t7}) \quad (y_5 > 0 \text{ for M7 to be saturated}).$$

Since SPICE produces values for V_{ds1} , etc., the data can be used to fit y_4 and y_5 without requiring new circuit simulations. The equations for y_4 and y_5 fit very well, with $R^2(R^2_{PRS}) = 1.000(1.000)$, and $0.996(0.957)$; they will be used in the yield optimization to constrain the operating regions of transistors M1 and M7.

The eleven points with poor gains ($y_4 < 0$ or $y_5 < 0$) are deleted from the dataset. The models for the gain, A_v , and the delay, T_2 , fit very well to the remaining 89 observations, with $R^2(R^2_{PRS}) = 0.999(0.982)$, and $0.998(0.969)$. The model for T_1 does not fit as well, however, with $R^2(R^2_{PRS}) = 0.976(0.559)$. We could try to improve the model for T_1 , but it turns out that the T_1 delays are small enough that they do not affect the yield anyway.

For a given C , the parametric yield can be predicted by Monte Carlo simulation using the fitted models, $\hat{y}_k(C + U, v)$. We generate 500 samples U_i , $i = 1, \dots, 500$ from the distribution $\Gamma(U)$. The predicted yield is

$$\hat{\Phi}(C) = \frac{1}{500} \sum_{i=1}^{500} \prod_{v=-1,0,1} I(\hat{y}_1(C + U_i, v), \dots, \hat{y}_5(C + U_i, v)). \quad (4.7)$$

The performance constraints must be met for all three levels of v , hence, the product of indicator functions.

The yield is optimized with respect to C using the simplex algorithm. This is computationally inexpensive because we are using the *fitted* models and do not have to run the simulator. For constraints $y_1 > 1000$, $y_2, y_3 < 170$ ns, and $y_4, y_5 > 0$, there are many designs with very high yield. This suggests that the performance constraints may be tightened. Re-optimizing with the new constraints on the two time delays, $y_2, y_3 < 130$ ns, produces the design C^* with $c_6 = 0.99$, $c_7 = -0.96$, $c_8 = -1.00$, and $c_9 = 0.77$. The estimated yield is $\hat{\Phi}(C^*) = 95\%$. Further tightening of this constraint leads to drastically lower yields. However, as the design C^* has the width M7 at its lower bound ($c_8 = -1.00$), experimentation with smaller widths of M7 may lead to further improvement. A similar comment may apply to c_6 and c_7 .

The regression models for the yield optimization may not approximate the performance variations around C^* very accurately. A confirmation experiment provides a more-accurate estimate of the yield. The confirmation experiment follows [3]. The assumed model of the responses is a quadratic of u_1, \dots, u_6 , and v . Data are collected according to the 48-run Latin hypercube design varying u_1, \dots, u_6 and v only. Models for y_1, \dots, y_5 have R^2 (R^2_{PRS}) = 1.000 (0.944), 0.995 (0.605), 1.000 (0.918), 1.000 (1.000), and 0.998 (0.928). This suggests the models of y_1, y_3, y_4 , and y_5 fit very well. The model for T_1 does not fit well, but again it does not matter. The predicted yield $\hat{\Phi}(C^*)$ obtained from 200 Monte Carlo samples with the fitted models is 93%. Figure 4.3 is a contour plot of the predicted parametric yield as the constraints on the gain and delays vary.

From running the circuit simulator at the same 200 Monte Carlo samples (this involves 600 circuit-simulator runs, because there are three levels of the operating condition v), the yield at C^* is 92%. This close agreement with 93% shows that the 48-run confirmation

T_1 and T_2

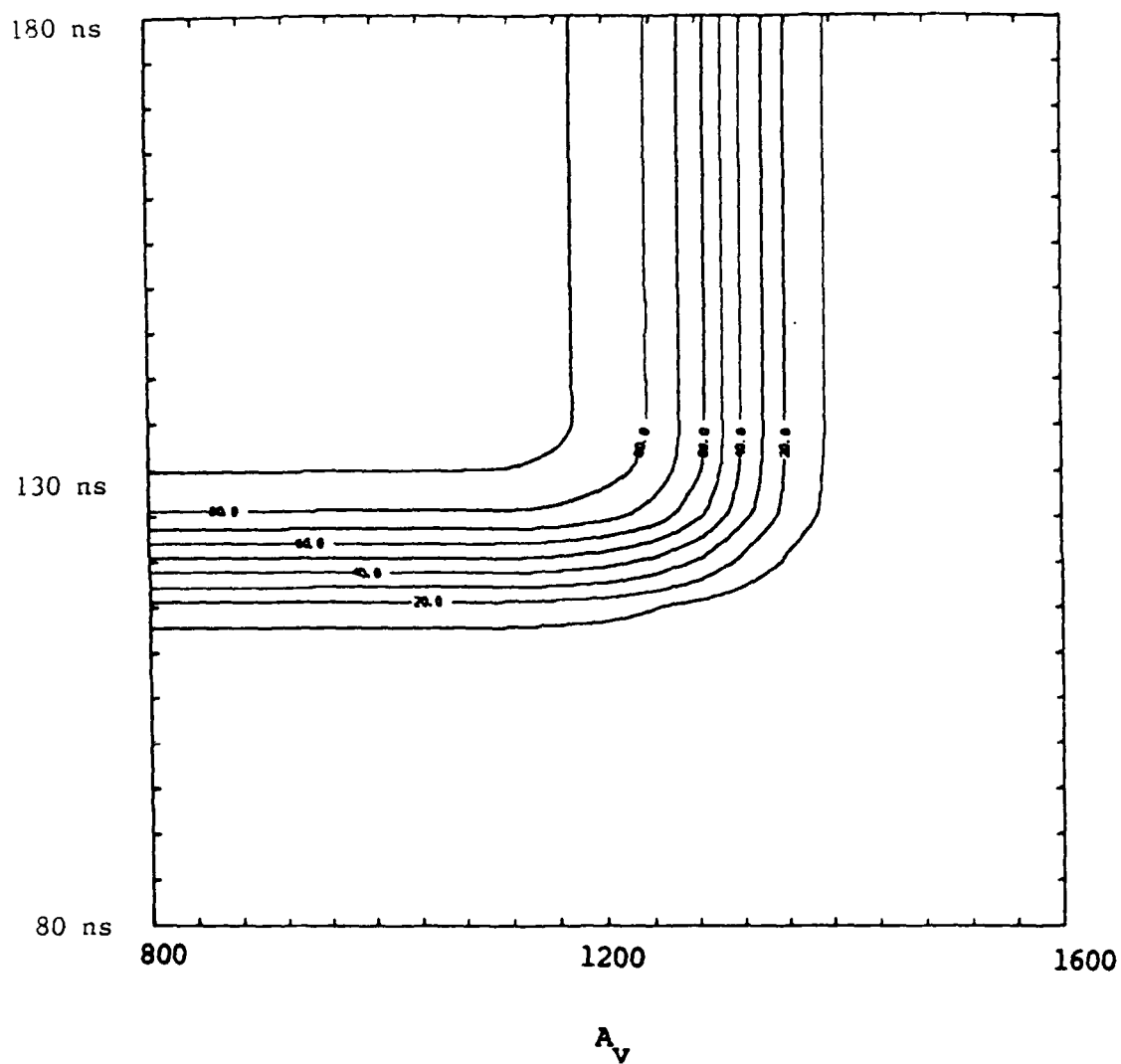


Figure 4.3 Contour plot of yield from 36-run confirmation experiment for Example A (yield contour interval = 10%).

experiment would have been sufficient. Moreover, the contour plot in Figure 4.4 of the yields from the SPICE runs agrees well with that in Figure 4.3.

Example B. CMOS Comparator II

A second CMOS comparator consisting of 53 MOS transistors is shown in Figure 4.5. The circuit topology follows from the techniques described in [38]. The performances of interest and their constraints are

$$A_v = \text{dc gain } (A_v > 30),$$

$$\omega_{3dB} = 3 \text{ dB bandwidth } (\omega_{3dB} > 60 \text{ MHz}).$$

In this example there are 12 parameters of interest. The first five are the uncontrollable statistical variations u_1, \dots, u_5 for the critical MOSFET parameters already described in Example A. The distributions of the device parameters are again independent Gaussian, but their $[-3\sigma, +3\sigma]$ ranges are different. The remaining seven parameters are designable, with no statistical variation:

$$c_6 = \text{width of transistors } M1A \text{ and } M1B$$

$$(30 \mu m \leq W(M1A, M1B) \leq 80 \mu m),$$

$$c_7 = \text{width of transistors } M2A \text{ and } M2B$$

$$(4 \mu m \leq W(M2A, M2B) \leq 12 \mu m),$$

$$c_8 = \text{ratio of the widths of } M3A \text{ and } M3B \text{ to } M2A \text{ and } M2B$$

$$(1.1 \leq W(M3A, M3B)/W(M2A, M2B) \leq 1.9),$$

T_1 and T_2

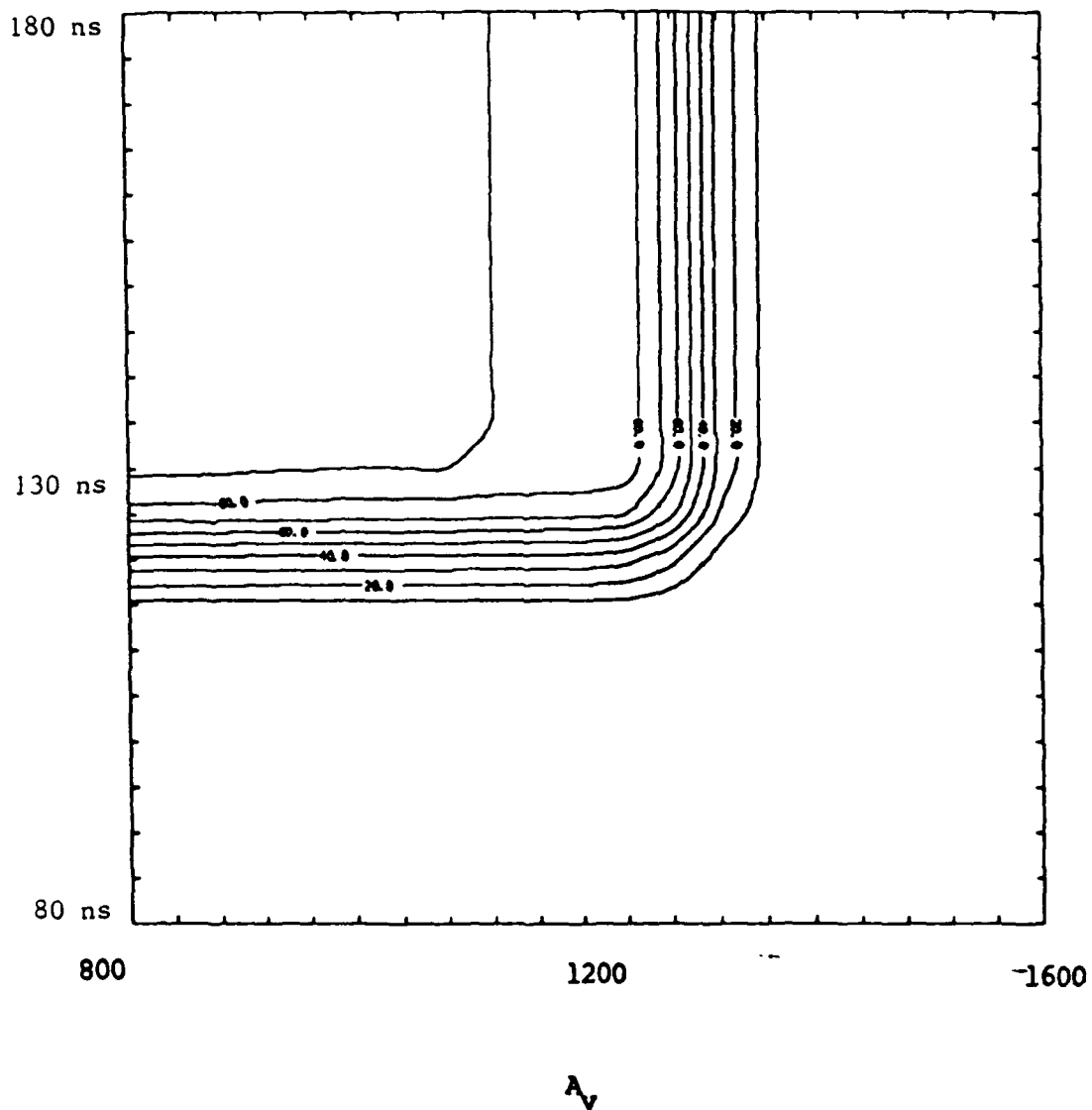


Figure 4.4 Contour plot of yield from 600-run circuit simulator runs for Example A (yield contour interval = 10%).

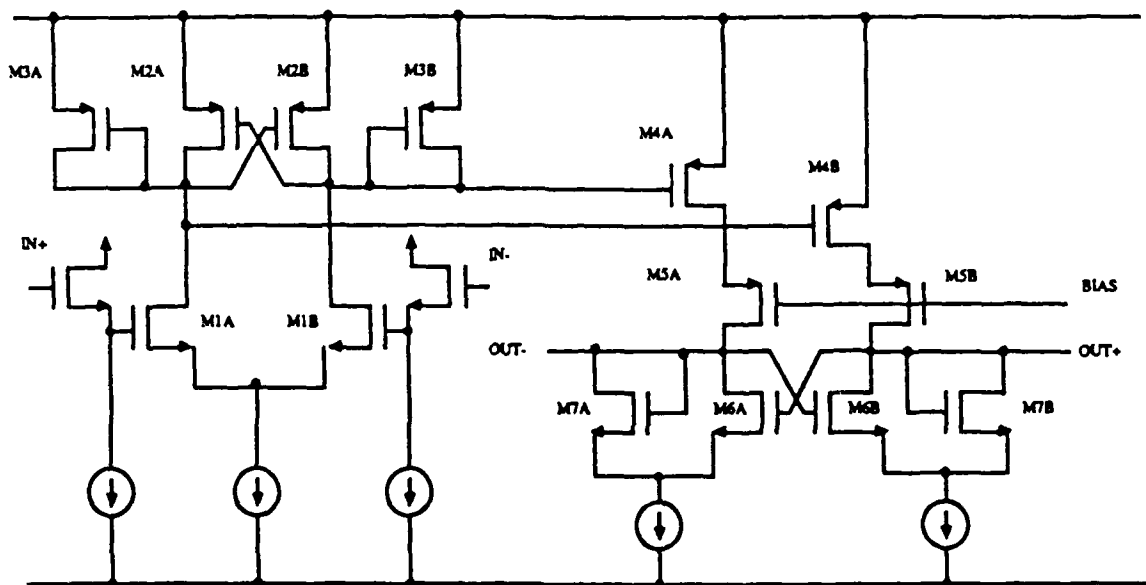


Figure 4.5 CMOS comparator for Example B.

c_9 = width of transistors $M4A$ and $M4B$

$$(15 \mu m \leq W(M4A, M4B) \leq 70 \mu m),$$

c_{10} = width of transistors $M5A$ and $M5B$

$$(15 \mu m \leq W(M5A, M5B) \leq 35 \mu m),$$

c_{11} = width of transistors $M6A$ and $M6B$

$$(4 \mu m \leq W(M6A, M6B) \leq 12 \mu m),$$

c_{12} = ratio of the widths of $M7A$ and $M7B$ to $M6A$ and $M6B$

$$(1.1 \leq W(M7A, M7B)/W(M6A, M6B) \leq 1.9).$$

All the parameters are normalized to the range $[-1, +1]$. Part of a 100-run Latin hypercube design for the 12 parameters $u_1, \dots, u_5, c_6, \dots, c_{12}$ is listed in Table 4.2, along with the data.

Modeling the gain by a quadratic model in $u_1, \dots, u_5, c_6, \dots, c_{12}$ gives a fairly poor predictor: $R^2 = 0.999$ but $R^2_{PRS} = 0.750$. To improve the prediction of the gain we note that the comparator circuit consists of two stages. If A_{v1} and A_{v2} denote the gains at the two stages, then

$$A_v = A_{v1} * A_{v2}. \quad (4.8)$$

We therefore try modeling A_{v1} and A_{v2} , from which A_v can be predicted.

The transistors in the first and second stages are connected together at nodes 1 and 2, and we use V_C to denote the common dc voltage at these two nodes, which is a parameter affecting A_{v2} . We assume that the models for

y_1 = dc gain of the first stage, A_{v1} ,

y_2 = dc voltage at nodes 1 and 2, V_C ,

Table 4.2 Part of 100-run Latin hypercube design for x_1, \dots, x_{12} and the observed y_1, \dots, y_4 for Example B.

| Run | x_1 | x_2 | x_3 | x_4 | x_5 | x_6 | x_7 | x_8 | x_9 | x_{10} | x_{11} | x_{12} |
|-----|-------|-------|-------|-------|-------|-------|-------|-------|-------|----------|----------|----------|
| 1 | -0.23 | 0.86 | -0.43 | 0.05 | 0.33 | 0.92 | 0.52 | 0.01 | -0.58 | 0.17 | 0.25 | -0.90 |
| 2 | 0.09 | 0.88 | -0.15 | -0.78 | -0.31 | -0.27 | -0.09 | 0.37 | 0.84 | 0.54 | 0.98 | 0.39 |
| 3 | 0.86 | 0.17 | -0.05 | -0.47 | 0.45 | -0.05 | -0.58 | -0.76 | 0.86 | -0.68 | 0.72 | -0.21 |
| 4 | 0.58 | -0.94 | 0.35 | 0.13 | -0.11 | 0.05 | 0.01 | 0.13 | 0.64 | -0.29 | -0.60 | 0.01 |
| 5 | -0.86 | -0.03 | 0.66 | 0.72 | 0.58 | -0.70 | -0.19 | 0.05 | 0.98 | 0.29 | 0.62 | 0.31 |
| . | | | | | | | | | | | | |
| 96 | -0.92 | 0.94 | 0.23 | 0.09 | -0.21 | 0.64 | -0.05 | -0.52 | -0.25 | -0.60 | -0.07 | -0.33 |
| 97 | 0.68 | 0.82 | 0.60 | -0.39 | 0.23 | -0.39 | 0.78 | -0.70 | -0.23 | 0.01 | 0.80 | 0.96 |
| 98 | 0.19 | -0.17 | 0.64 | -0.07 | -0.94 | -0.23 | -0.33 | -0.82 | 0.15 | 0.05 | -0.70 | 0.37 |
| 99 | 0.60 | 0.15 | 0.15 | 0.11 | 0.17 | -0.58 | 0.49 | -0.66 | 0.29 | -0.03 | -0.13 | 0.43 |
| 100 | -0.52 | 0.41 | 0.92 | -1.00 | -0.01 | 0.33 | -0.03 | 0.43 | -0.60 | 0.03 | -0.47 | 0.03 |

| Run | y_1 | y_2 | y_3 | y_4 |
|-----|-------|-------|-------|-------|
| 1 | 7.1 | 7.3 | 0.93 | 43.0 |
| 2 | 5.0 | 3.8 | 0.81 | 68.1 |
| 3 | 8.6 | 5.2 | 0.56 | 46.4 |
| 4 | 5.8 | 5.5 | 0.50 | 63.0 |
| 5 | 5.8 | 3.7 | 0.63 | 50.1 |
| . | | | | |
| 96 | 9.5 | 2.6 | 0.80 | 54.1 |
| 97 | 7.6 | 2.4 | 0.92 | 68.0 |
| 98 | 9.4 | 4.0 | 0.55 | 54.1 |
| 99 | 7.1 | 3.9 | 0.79 | 54.1 |
| 100 | 5.8 | 3.3 | 0.58 | 92.5 |

should be functions of u_1, \dots, u_5 and c_6, \dots, c_9 , while

y_3 = dc gain of the second stage, A_{v2} ,

depends on $u_1, \dots, u_5, c_9, \dots, c_{12}$, and V_C .

In the models for y_3 below, y_2 is treated as another input parameter, and we write (x_1, \dots, x_{13}) for $(u_1, \dots, u_5, c_6, \dots, c_{12}, y_2)$. The assumed quadratic equations for A_{v1} and V_C are

$$y_k = \beta_{0k} + \sum_{i=1}^9 \beta_{ik} x_i + \sum_{i=1}^9 \sum_{j=i}^9 \beta_{ijk} x_i x_j + \text{error}, \quad k = 1, 2. \quad (4.9)$$

Similarly, the assumed model for A_{v2} is a quadratic in $(x_1, \dots, x_5, x_9, x_{10}, x_{11}, x_{12}, x_{13})$. The regression models of y_1, \dots, y_3 have R^2 (R^2_{PRS}) values of 0.999 (0.995), 1.000 (0.999), and 0.995 (0.945), respectively, suggesting very good predictive capabilities for A_{v1} , V_C , and A_{v2} . We note that A_{v1} is strongly affected by c_6 and c_8 , while A_{v2} depends strongly on c_{12} .

Modeling

y_4 = the 3 dB bandwidth, ω_{3dB} ,

by a quadratic model in $u_1, \dots, u_5, c_6, \dots, c_{12}$ gives a fairly good predictor: $R^2 = 0.999$ and $R^2_{PRS} = 0.851$. Attempting to improve the predictor by decomposing the circuit into two stages is not successful for bandwidth. An approximation relating the overall bandwidth to the bandwidths of the two stages, ignoring higher-order poles, leads to substantial error here. Thus, we retain the quadratic model in $u_1, \dots, u_5, c_6, \dots, c_{12}$ for the bandwidth.

Monte Carlo estimates of yield for a circuit defined by C are based on 500 samples, U_i , $i = 1, \dots, 500$. The fitted models are used to make predictions of the bandwidth, the gain of stage 1, and the voltage V_C . Then, the predicted voltages are fed into the model \hat{y}_3 to predict

the gain of stage 2. The predicted gain is computed from Eq. (4.8). The estimated yield $\hat{\Phi}(C)$ is the percentage of samples with $A_v > 30$, and $\omega_{3dB} > 60$ MHz.

The yield is optimized with respect to C using the simplex algorithm. A design C^* with $c_6 = 0.60$, $c_7 = -1.00$, $c_8 = 0.10$, $c_9 = -1.00$, $c_{10} = 0.85$, $c_{11} = -0.98$, $c_{12} = -0.58$, and $\hat{\Phi}(C^*) = 100\%$ is found.

For the confirmation experiment, we use a 36-run Latin hypercube design varying only the uncontrollable parameters u_1, \dots, u_5 . The models for y_1, \dots, y_4 have R^2 (R^2_{PRS}) = 1.000 (1.000), 1.000 (1.000), 1.000 (0.998), and 0.945 (0.649), respectively. This suggests that the models (\hat{y}_1 , \hat{y}_2 , and \hat{y}_3) of A_{v1} , A_{v2} , and V_C fit very well. The predictor \hat{y}_4 of ω_{3dB} is less accurate. Nonetheless, the yield of 99.0% estimated by the four models from 500 Monte Carlo samples agrees well with a predicted yield of 99.2% from running the circuit simulator 500 times from the same Monte Carlo samples. Again, the small confirmation experiment is adequate. Figure 4.6 is a contour plot of the yield (from the 500 circuit simulator runs) as the constraints on the gain and bandwidth vary. It indicates that there is a large region where the yield is close to 100%. This explains why our predictor of bandwidth is adequate here. The plot also indicates that the bandwidth is not a smooth function of the inputs: there is a rapid decrease in yield in certain parts of the region. This explains the difficulty in modeling the bandwidth by a quadratic function. The more-flexible models suggested in [35] may lead to a better predictor.

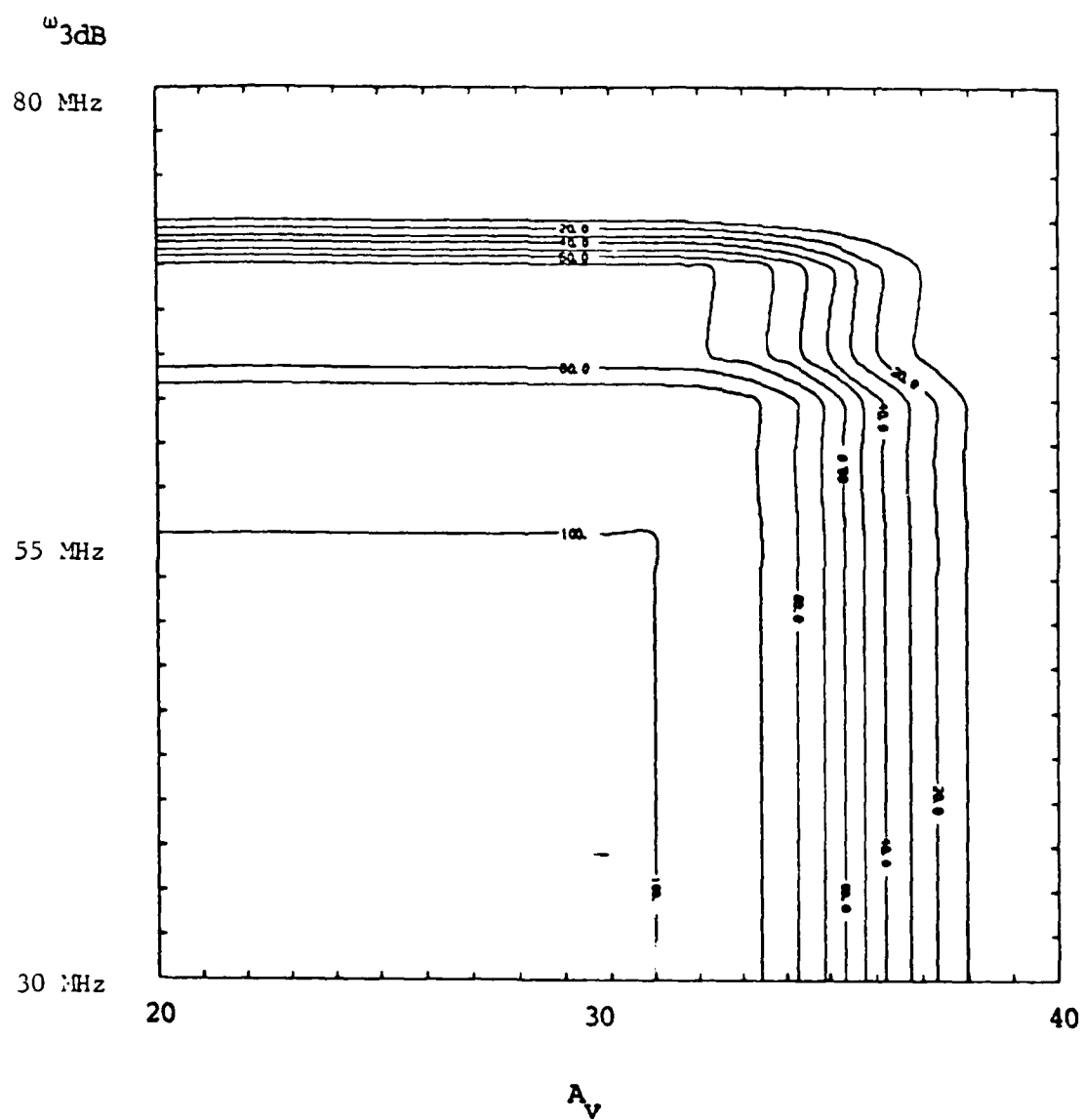


Figure 4.6 Contour plot of yield from 500-run circuit simulator runs for Example B (yield contour interval = 10%).

4.3. Concluding Remarks

In this chapter we presented a new method for parametric yield optimization of MOS VLSI circuits. By constructing computationally cheap models for the circuit performances, we achieve high yields with relatively few simulator runs. Based on our experience on this and other examples:

1. Engineering knowledge should be used to select the most basic performances, as illustrated by the modeling of A_{v1} and A_{v2} in the second example. These are more likely to admit simple polynomial approximations of high accuracy, from which the performances of interest can be constructed.
2. If there are ill-behaved observations, such as catastrophic failures, and an assumed model does not fit well, the designer should identify the causes of the failures. As we demonstrated in the first example, additional design constraints can be included to keep the yield optimization away from undesirable regions of the designable parameters.
3. Points 1 and 2 suggest that statistical modeling should incorporate engineering insights.
4. With careful modeling, accurate approximations to crucial performances can be developed with relatively few circuit simulations. Effective yield optimization is possible with these computationally inexpensive models.

CHAPTER 5.

PARAMETER DESIGN METHOD FOR OFF-LINE QUALITY CONTROL

Taguchi's off line quality control method [9] for product and process improvements has gained much interest recently in engineering. Instead of maximizing the parametric yield, the Taguchi approach [9] employs the design and analysis of experiments to design quality "into" products, such that they are insensitive to the manufacturing process and environmental variations. Taguchi's approach often requires a prohibitively large number of experimental runs, however. The Taguchi method collapses data from many (circuit simulator) runs into his so called "signal-to-noise ratio," just as the yield gradient method collapses data when computing the yield gradient.

In this chapter we extend the yield optimization method in Chapter 4 to achieve off-line quality control. Again, the proposed method models the circuit performance as a function of all parameters of interest. For a given set of designable parameters the fitted models predict the circuit performances as the uncontrollable parameters vary. This leads to a prediction of the Taguchi loss statistic, instead of parametric yield. The loss statistic is then numerically minimized with respect to the designable parameters. We give a circuit example where the Taguchi objectives are met with about two-thirds fewer runs than [9].

The Taguchi method is described in Section 5.1. The proposed method is detailed in Section 5.2. Section 5.3 compares the Taguchi and the proposed approach in a CMOS clock skew minimization example. Discussion is given in Section 5.4.

5.1. Taguchi's Parameter Design Method

Taguchi's off-line quality control methods [9] emphasize designing quality into products (circuits), so that they are less sensitive to sources of variability. Parameter design, an important step in off-line quality control, is the search for levels of designable (controllable) parameters that lead to a product robust to the variability in the manufacturing process and environmental conditions (noise). A key feature in [9] is the separation of the designable parameters and uncontrollable statistical variation for the design of experiment. Kackar [39], and Hunter [40] gave very readable accounts of the main ideas; Kackar and Shoemaker [41], and Phadke [42] provided some examples.

We use y to denote the circuit performance of interest, and C and U to denote the controllable and uncontrollable variations, respectively. The distribution of U is denoted by $\Gamma(U)$. Taguchi measures the quality of a circuit C by a *loss statistic* [9]:

$$L(C) = \int (y(C,U) - y_{target})^2 \Gamma(U) dU, \quad (5.1)$$

which is the expected squared deviation of the circuit performance from its target value, y_{target} . The objective of parameter design is to find C that minimizes $L(C)$.

Taguchi optimizes $L(C)$ with respect to C in a two-step procedure [40]. First, signal-to-noise ratios are computed from the data, which are then used to optimize the loss statistic. The connection between the signal-to-noise ratio and the loss statistic is given in [43].

Taguchi's design strategy for off-line quality control involves six steps. The example will illustrate details of its implementation.

STEP 1: Assume a model of the loss statistic

$$L(C) = g(C, \beta) + \text{error}. \quad (5.2)$$

The model g is an assumed-known function of C (for example, a polynomial model), and β is a vector of unknown constants to be estimated from the data. The error term represents systematic departure from the assumed model. We prefer to minimize $L(C)$, because it underlies the Taguchi philosophy, rather than maximize a signal-to-noise ratio. The disadvantages of the signal-to-noise ratio are discussed in [44].

STEP 2: Design a control array experiment for C and a noise array experiment for U . For every C in the control array, simulate the circuits at every U in the noise array.

STEP 3: Compute estimates of the loss statistic $\hat{L}(C)$ for every point in the control array. Fit the data to model (5.2).

STEP 4: Minimize $\hat{L}(C)$ with respect to C , subject to design constraints.

STEP 5: Conduct a confirmatory experiment to evaluate $L(C)$ by fixing C at the value(s) found in STEP 4 and varying U according to $\Gamma(U)$.

A major problem of Taguchi's approach is the large number of experimental runs due to the crossing of the control and noise arrays. Moreover, it is often difficult to identify an appropriate model for the loss function. In the following section, we will discuss how the

yield optimization method in Chapter 5 is adapted to achieve off-line quality control, but with far fewer runs than the Taguchi experiment.

5.2. Proposed Off-line Quality Control Method

The proposed approach involves five steps:

STEP 1: Design a single experiment to predict the performance as a function of the controllable parameters C and the uncontrollable statistical variables U :

$$y(C, U) = f(C, U, \beta) + \text{error}. \quad (5.3)$$

The model f is an assumed-known function of both C and U , and β is a vector of unknown constants. We do not use separate control and noise arrays; considerable economy in the number of observations can result from designing a single experiment for both factors.

STEP 2: Simulate the circuits at the design points, and fit the performance model $\hat{y}(C, U)$. For a given C predict the loss statistic $L(C)$ from the estimated response function. For example, the predicted loss statistic (5.2) is

$$\hat{L}(C) = \int \hat{y}^2(C, U) \Gamma(U) dU. \quad (5.4)$$

In the example below, the density $\Gamma(U)$ is taken as combinations of the best- and worst-case current files of the transistors. This is mainly for convenience.

STEP 3: Minimize $\hat{L}(C)$ as a function of C . In practice, the mathematical optimization will be tempered by engineering and cost considerations.

STEP 4: Conduct a confirmatory experiment to evaluate $L(C)$ by fixing C at the value(s) found in STEP 3 and varying U according to $\Gamma(U)$.

STEP 5: Iterate if necessary. For instance, if the optimization step suggests values of C outside the design region which are technically and economically feasible, then the design region could be shifted.

5.3. Example: Minimization of Process Dependent Clock Skew

This example considers a CMOS clock driver [45] shown in Figure 5.1. From the master clock CLK_M , the circuit generates outputs CLK and \overline{CLK} . The clock skew is defined to be the difference in the output delay times of the CLK and \overline{CLK} signals. Because each clock signal switches twice per machine cycle, two clock skews S_R and S_F can be measured (in units of nanoseconds), as illustrated in Figure 5.2.

The design objective is to determine the channel widths $\{w_1, \dots, w_6\}$ for the transistors M_1, \dots, M_6 in Figure 5.1 that give the smallest clock skew in the presence of device parameter variability. To allow for quadratic effects, the experiment is carried out with each width at three levels, denoted by -1, 0, 1.

Shoji [45] investigated the clock skew example and proposed an empirical method for minimizing the skew. In [45], the uncontrollable statistical variabilities are represented by the high (H), medium (M), and low (L) current files of the P- and N-channel MOSFET transistors. These device parameter combinations are coded 1,...,5 below. For the purpose of comparison, we will use these device parameter combinations in our experiment.

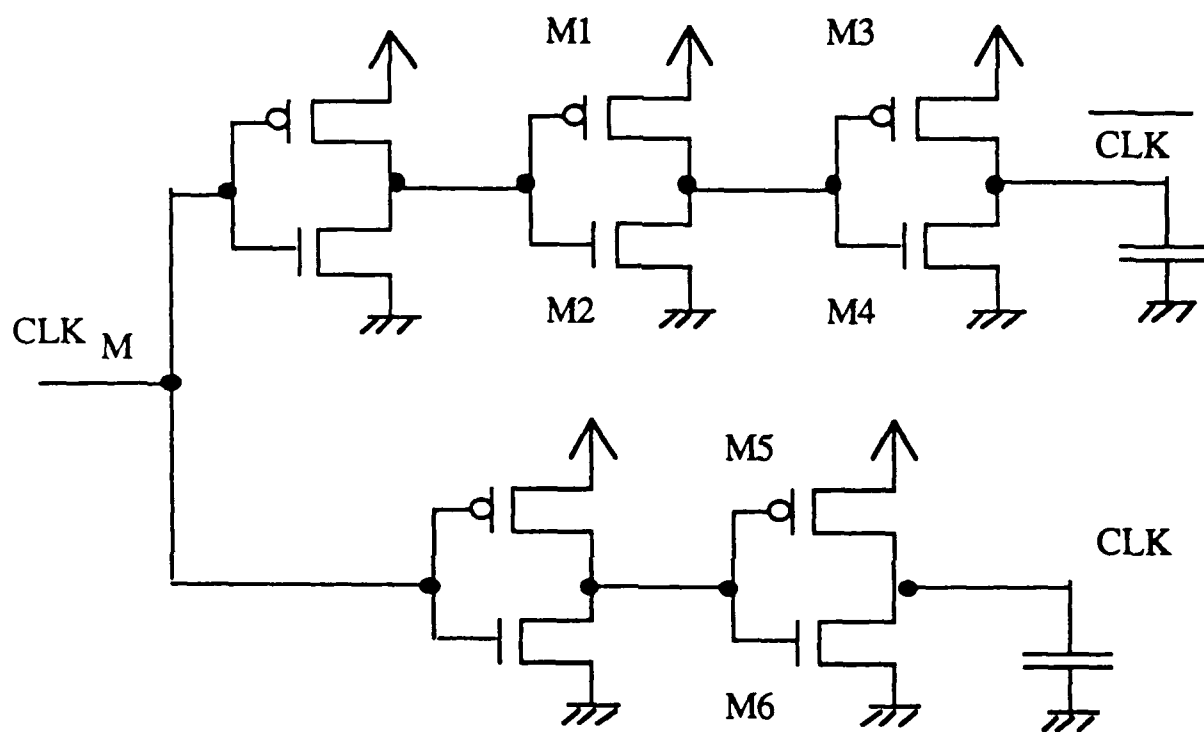


Figure 5.1 CMOS clock driver circuit.

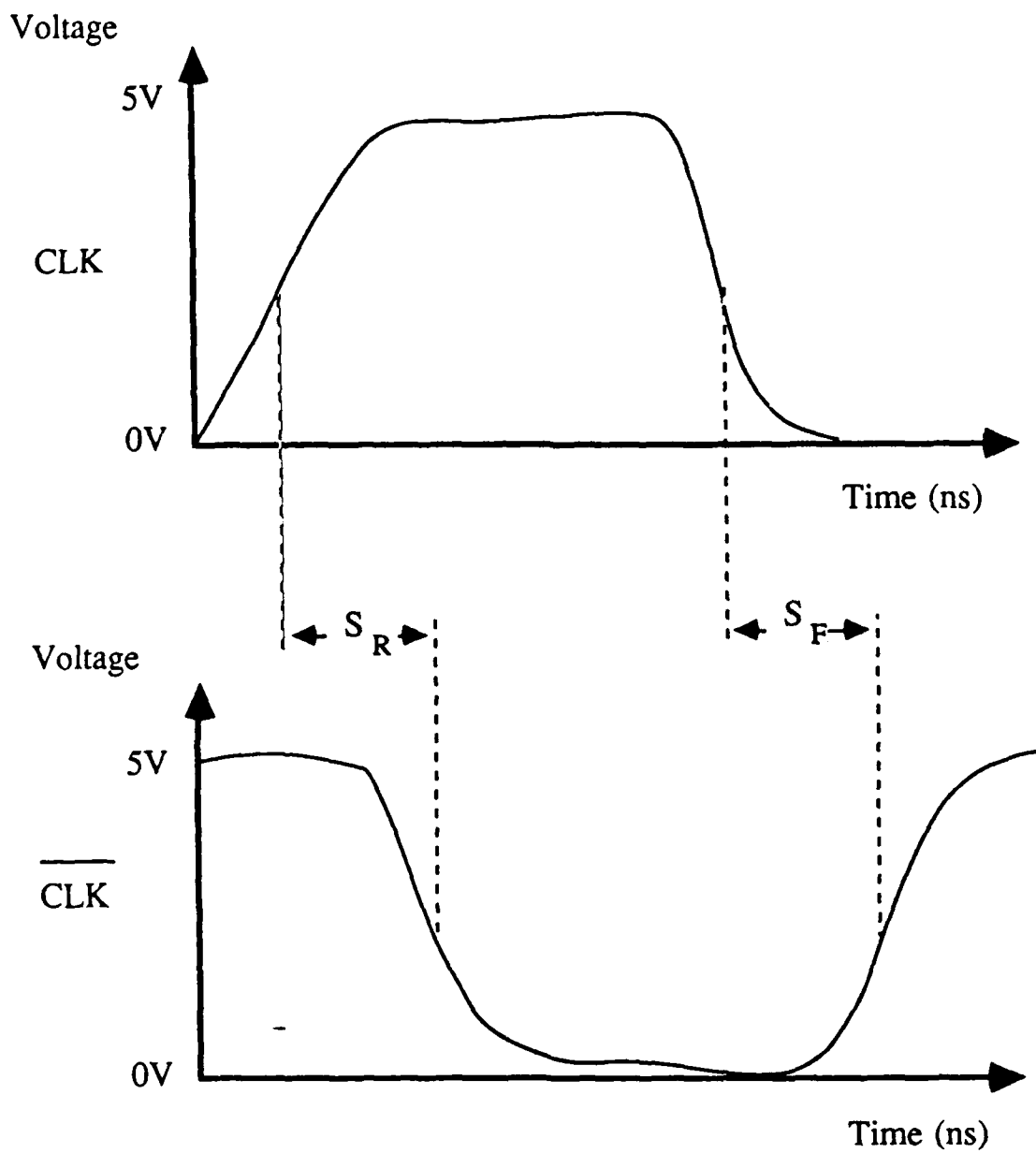


Figure 5.2 CMOS clock skews.

We will consider loss statistics that combine two skews. For fixed $w = (w_1, \dots, w_6)$, there are five pairs of skews corresponding to device parameters {PH-NH, PH-NL, PM-NM, PL-NH, PL-NL}. Denote these skews by $S_1(w), \dots, S_{10}(w)$. The target skew is zero, and the logarithm of the average squared-error loss is

$$L_{sq}(w) = \log \left[\frac{1}{10} \sum_{i=1}^{10} S_i(w) \right]. \quad (5.5)$$

The logarithmic transformation, being monotonic, does not affect the optimal w but is relevant in Section 5.3.2 when L_{sq} is modeled directly. A more-conservative performance measure is the worst-case skew

$$L_{wc}(w) = \max [|S_1(w)|, \dots, |S_{10}(w)|]. \quad (5.6)$$

5.3.1. Modeling the circuit performances

Several considerations determined the choice of model for the clock skews as functions of the controllable parameters and the uncontrollable statistical variation.

- Because curvature and interactions cannot be ruled out we adopt a second-order polynomial model.
- There are only five combinations of the uncontrollable statistical variations {PH-NH, ..., PL-NL}. For simplicity, therefore, we treat these combinations as a single qualitative factor ζ at five levels 1, ..., 5.

- There are two chains of transistors in Figure 5.1: (M1,...,M4) and (M5,M6). This suggests that only seven interactions w_1w_2 , w_1w_3 , w_1w_4 , w_2w_3 , w_2w_4 , w_3w_4 , and w_5w_6 involving transistors in the same chain need be considered.
- No similar reasoning leads to a reduction in the number of interactions between the widths and the device factor, however. Although one might suspect, for example, that there would be no interaction between the widths of the P-channel transistors and the component of the device factor representing variability in current-driving capabilities of the N-channel transistors, this is not the case. The two types of transistors are interconnected. The interconnections between the transistors allow the effects of the device parameters to differ at various transistor sizes. Easterling [46] pointed out this connection between interactions and robustness to sources of variability.

Thus, we model the two skews as a function of C and $\zeta = j$ by

$$\begin{aligned}
 Y(w, j) = & \beta_0 + \beta_1w_1 + \dots + \beta_6w_6 + \beta_{11}w_1^2 + \dots + \beta_{66}w_6^2 \\
 & + \beta_{12}w_1w_2 + \beta_{13}w_1w_3 + \beta_{14}w_1w_4 + \beta_{23}w_2w_3 + \beta_{24}w_2w_4 \\
 & + \beta_{34}w_3w_4 + \beta_{56}w_5w_6 + \gamma_j + \delta_{1j}w_1 + \dots + \delta_{6j}w_6 + \epsilon.
 \end{aligned} \tag{5.7}$$

The unknown constants $\gamma_1, \dots, \gamma_5$ and $\delta_{11}, \dots, \delta_{65}$ are the main effects for the qualitative device factor and the interaction effects between the designable and device factors. Because the observations are derived from a deterministic circuit simulator, there is no random error, and ϵ represents systematic departure from the assumed linear model. Because not all the unknown constants are identifiable, we arbitrarily set $\gamma_5 = \delta_{15} = \dots = \delta_{65} = 0$. This leaves 48 unknown constants to be estimated.

We designed a 60-observation experiment to estimate the 48 unknown constants in model (5.7). The number of runs is fairly arbitrary, clearly, at least 48 are needed, and we wanted a modest number of degrees of freedom to assess potential lack of fit. Again, the average mean-squared error criterion in the ACED package [28] was used to obtain the design (see Section A.1). (This criterion also includes the variance arising from random error; we weighted the bias component to be dominant because there is no random error) The use of a computer package such as ACED also circumvents the difficulties in designing this experiment: only some of the interactions need to be estimated and the five-level device factor space is not a regular factorial arrangement. The experimental design and the resulting data are given in Table 5.1.

The two skews are separately modeled via Eq. (5.7). Least squares estimation of the unknown constants allows us to predict the two skews at untried levels of the designable and device factors. In the presence of systematic error rather than random error, statistical testing is inappropriate. Nonetheless, the root mean squared errors of the least squares analyses for the two skews are 0.03 and 0.08 (relative to data ranges of about -3.9 to 0.2 and -2.2 to 3.8). The R^2 values are 1.000 and 0.999, suggesting that the models fit well. We note that

- The first-order effects for both the designable and device factors are all large.
- Many of the second-order (quadratic and interaction) effects are moderately large.
- The contrast between high and low levels of the N-channel device factor is larger than that for P. It is often found that the N-channel variability is more critical.

Table 5.1 Experimental design and data for modeling clock skews.
(Two skews are observed for each run.)

| Run | Transistor Width w | | | | | | Noise Level | Skews | |
|-----|----------------------|----|----|----|----|----|-------------|--------|--------|
| 1 | -1 | -1 | -1 | -1 | -1 | -1 | 3 | -1.289 | -0.307 |
| 2 | -1 | -1 | -1 | -1 | -1 | 0 | 4 | -0.636 | -1.199 |
| 3 | -1 | -1 | -1 | -1 | 1 | -1 | 1 | -1.219 | 0.907 |
| 4 | -1 | -1 | -1 | 1 | 1 | -1 | 2 | -1.151 | 1.678 |
| 5 | -1 | -1 | -1 | 1 | 1 | 1 | 3 | -0.449 | -0.422 |
| 6 | -1 | -1 | -1 | 1 | 1 | 1 | 5 | -0.510 | -0.343 |
| 7 | -1 | -1 | 0 | -1 | -1 | -1 | 5 | -2.758 | 0.157 |
| 8 | -1 | -1 | 1 | -1 | -1 | 1 | 2 | -2.414 | -1.309 |
| 9 | -1 | -1 | 1 | 0 | -1 | 1 | 5 | -1.920 | -1.633 |
| 10 | -1 | -1 | 1 | 1 | -1 | 1 | 4 | -0.809 | -1.546 |
| 11 | -1 | -1 | 1 | 1 | 0 | 0 | 1 | -1.227 | -0.496 |
| 12 | -1 | 0 | -1 | -1 | 0 | 1 | 2 | -1.412 | 0.041 |
| 13 | -1 | 0 | -1 | 1 | 0 | 1 | 4 | -0.452 | -0.628 |
| 14 | -1 | 0 | 0 | 0 | -1 | -1 | 1 | -1.127 | 0.062 |
| 15 | -1 | 0 | 1 | -1 | 1 | 0 | 5 | -3.860 | 2.011 |
| 16 | -1 | 0 | 1 | 0 | 0 | -1 | 4 | -2.107 | 0.862 |
| 17 | -1 | 1 | -1 | -1 | -1 | -1 | 2 | -2.300 | 1.350 |
| 18 | -1 | 1 | -1 | -1 | -1 | 1 | 3 | -1.118 | -0.466 |
| 19 | -1 | 1 | -1 | 0 | -1 | 0 | 5 | -1.495 | 0.070 |
| 20 | -1 | 1 | -1 | 1 | 0 | 1 | 1 | -0.512 | -0.236 |
| 21 | -1 | 1 | -1 | 1 | 1 | -1 | 4 | -1.184 | 1.592 |
| 22 | -1 | 1 | 1 | -1 | -1 | -1 | 1 | -2.126 | 0.479 |
| 23 | -1 | 1 | 1 | -1 | 1 | 1 | 4 | -2.504 | 0.931 |
| 24 | -1 | 1 | 1 | 0 | 1 | -1 | 3 | -2.769 | 2.567 |
| 25 | -1 | 1 | 1 | 1 | 1 | -1 | 5 | -3.315 | 3.759 |
| 26 | -1 | 1 | 1 | 1 | 1 | 1 | 2 | -1.982 | 1.149 |
| 27 | 0 | -1 | -1 | -1 | 0 | -1 | 5 | -1.927 | 0.365 |
| 28 | 0 | -1 | -1 | 0 | 1 | 1 | 4 | -0.452 | -0.922 |
| 29 | 0 | -1 | 0 | 1 | -1 | 0 | 5 | -0.855 | -2.175 |
| 30 | 0 | -1 | 1 | -1 | 0 | 0 | 4 | -1.768 | -0.748 |

(continued)

Table 5.1 (continued)

| | | | | | | | | | |
|----|---|----|----|----|----|----|---|--------|--------|
| 31 | 0 | -1 | 1 | 1 | -1 | 1 | 3 | -0.715 | -2.177 |
| 32 | 0 | 0 | -1 | 0 | 1 | 1 | 1 | -0.510 | -0.283 |
| 33 | 0 | 0 | 0 | 0 | -1 | 0 | 2 | -1.401 | -0.715 |
| 34 | 0 | 0 | 1 | 1 | 0 | 1 | 5 | -1.576 | -0.639 |
| 35 | 0 | 1 | -1 | -1 | 1 | -1 | 3 | -1.841 | 2.004 |
| 36 | 0 | 1 | 0 | 0 | 0 | 1 | 5 | -1.704 | 0.004 |
| 37 | 0 | 1 | 1 | 1 | 1 | 0 | 4 | -1.491 | 0.724 |
| 38 | 1 | -1 | -1 | -1 | 1 | 1 | 2 | -0.852 | -1.439 |
| 39 | 1 | -1 | -1 | 0 | 0 | 1 | 3 | -0.266 | -2.196 |
| 40 | 1 | -1 | -1 | 1 | -1 | -1 | 1 | -0.067 | -1.691 |
| 41 | 1 | -1 | 0 | -1 | 1 | 0 | 1 | -1.155 | -0.583 |
| 42 | 1 | -1 | 0 | 1 | 1 | -1 | 3 | -1.096 | 0.053 |
| 43 | 1 | -1 | 0 | 1 | 1 | 0 | 4 | -0.649 | -1.010 |
| 44 | 1 | -1 | 1 | -1 | -1 | 1 | 1 | -0.958 | -1.978 |
| 45 | 1 | -1 | 1 | 0 | 1 | 0 | 5 | -2.049 | -0.687 |
| 46 | 1 | -1 | 1 | 1 | -1 | -1 | 2 | -1.545 | -1.504 |
| 47 | 1 | 0 | -1 | -1 | -1 | 1 | 5 | -0.928 | -2.081 |
| 48 | 1 | 0 | 0 | -1 | 1 | 1 | 4 | -1.197 | -0.423 |
| 49 | 1 | 0 | 1 | -1 | 1 | 1 | 3 | -2.007 | -0.161 |
| 50 | 1 | 1 | -1 | -1 | 0 | 0 | 1 | -0.889 | -0.390 |
| 51 | 1 | 1 | -1 | 1 | -1 | -1 | 3 | -0.659 | -0.482 |
| 52 | 1 | 1 | -1 | 1 | -1 | -1 | 5 | -1.166 | -0.121 |
| 53 | 1 | 1 | -1 | 1 | -1 | 1 | 2 | -0.285 | -1.440 |
| 54 | 1 | 1 | -1 | 1 | -1 | 1 | 4 | 0.229 | -1.841 |
| 55 | 1 | 1 | 0 | -1 | 1 | 1 | 5 | -2.269 | 0.539 |
| 56 | 1 | 1 | 0 | 1 | -1 | 1 | 1 | -0.236 | -1.312 |
| 57 | 1 | 1 | 0 | 1 | -1 | 1 | 1 | -1.642 | -0.343 |
| 58 | 1 | 1 | 1 | -1 | -1 | 0 | 5 | -3.055 | -0.533 |
| 59 | 1 | 1 | 1 | -1 | 1 | -1 | 2 | -3.522 | 2.585 |
| 60 | 1 | 1 | 1 | 1 | 0 | 0 | 3 | -1.440 | -0.175 |

For fixed w , the two models are used to predict five pairs of skews corresponding to the five device factor levels. Denote these 10 predictions by $\hat{S}_1(w), \dots, \hat{S}_{10}(w)$. The loss statistics in Eqs. (5.5) and (5.6) can then be estimated by

$$\hat{L}_{sq}(w) = \log \left[\frac{1}{10} \sum_{i=1}^{10} \hat{S}_i^2(w) \right] \quad (5.8)$$

and

$$\hat{L}_{wc}(w) = \max[|\hat{S}_1(w)|, \dots, |\hat{S}_{10}(w)|]. \quad (5.9)$$

The next step is to minimize either of these loss statistics with respect to w . Discussion of this and the validation from confirmatory experiments will be presented in Section 5.3.3, along with a comparison of alternative design and modeling strategies.

5.3.2. Modeling the loss statistics directly

For comparison, we also conducted an experiment with separate control and noise arrays, as has been advocated by Taguchi for optimizing through direct modeling of a performance measure.

The choice of a model for a loss statistic L is problematic. Whereas the engineer may have substantial background knowledge concerning the underlying circuit performance, approximate models for complex loss statistics are typically not so intuitive. In this example, when modeling the skews, there is an engineering basis for omitting some designable-factor interactions. However, this need not imply that the same interactions are negligible when modeling the loss statistics, which are nonlinear functions of the skews. Indeed, these interactions turn out to have fairly large effects. As a simple illustration, $Y = x_1 + x_2$ has no

interaction between two factors x_1 and x_2 , but the loss Y^2 , for example, clearly does. A log transformation is often suggested to reduce interaction effects in signal-to-noise ratios similar to the squared error loss in Eq. (5.8) (see for example [39]). In the absence of engineering intuition, however, we adopt a full second-order model in w for both loss statistics (though hesitantly for the non-smooth L_{wc}):

$$L(w) = \beta_0 + \sum_{i=1}^6 \beta_i w_i + \sum_{i=1}^6 \sum_{j=1}^6 \beta_{ij} w_i w_j + \epsilon. \quad (5.10)$$

There are 28 unknown constants, and we designed a 40-run (control array) experiment for the designable factors, again using ACED. As when modeling the clock skews, the size of the control array is somewhat arbitrary. A minimum of 28 runs is needed, and we allow 12 extra degrees of freedom to measure the lack of fit. Crossing the control array with the noise array of size five leads to a total of 200 runs. Part of the experimental design and data are given in Table 5.2. For a given w in the control array, the data generated across the noise array collapsed to the loss statistics $L_{sq}(w)$ and $L_{wc}(w)$ in Eqs. (5.8) and (5.9). Fitting model (5.10) to these observed statistics provides direct predictions to $L_{sq}(w)$ and $L_{wc}(w)$ at untried w 's, which can be optimized with respect to w . The root mean squared errors for the L_{sq} and L_{wc} fits are 0.1 and 0.25 relative to data ranges of -0.5 to 0.8 and 1.0 to 3.9. Fitting by least squares gives R^2 values of 0.824 for L_{sq} in Eq. (5.8) and 0.907 for L_{wc} in Eq. (5.9). Qualitatively, then, the models do not fit quite as well as those for the skews.

As noted in Section 5.1, this experiment does not strictly follow the pattern of analysis in the examples given by Taguchi [9]. That paradigm fits additive models to the loss statistics (ignoring interactions) and would optimize the level of each transistor width separately. In this example, using the data from a (40×5) -run experiment, such an approach leads to

Table 5.2 Part of the crossed array experimental design and data for modeling loss statistics directly. (Two skews are observed for each run.)

| Runs | Transistor Widths w | | | | | | Skew at Noise Level | | | | |
|-------|-----------------------|----|----|----|----|----|---------------------|-------|-------|-------|-------|
| | | | | | | | 1 | 2 | 3 | 4 | 5 |
| 1-5 | -1 | -1 | -1 | -1 | 0 | 1 | -0.73 | -1.12 | -1.01 | -0.83 | -1.30 |
| | | | | | | | -0.76 | -0.78 | -0.76 | -0.86 | -0.78 |
| 6-10 | -1 | -1 | -1 | 1 | 1 | -1 | -0.74 | -1.15 | -1.01 | -0.88 | -1.43 |
| | | | | | | | 0.60 | 1.68 | 1.16 | 0.72 | 1.97 |
| 11-15 | -1 | -1 | 0 | 0 | -1 | -1 | -0.98 | -1.85 | -1.44 | -1.03 | -2.06 |
| | | | | | | | -0.46 | 0.01 | -0.34 | -0.57 | -0.13 |
| 16-20 | -1 | -1 | 1 | -1 | 0 | -1 | -2.11 | -3.45 | -2.85 | -2.39 | -3.91 |
| | | | | | | | 0.39 | 1.22 | 0.79 | 0.47 | 1.43 |
| 21-25 | -1 | -1 | 1 | 1 | 1 | 1 | -1.15 | -1.44 | -1.46 | -1.39 | 1.79 |
| | | | | | | | -0.29 | -0.29 | -0.19 | -0.22 | -0.09 |
| : | | | | | | | | | | | |
| : | | | | | | | | | | | |
| : | | | | | | | | | | | |

extremely poor predictions (given below). Also, Taguchi's L_{18} experimental design could be used for the designable factors. The comparison we make is more consistent with our methods.

5.3.3. Results and comparisons

We now give results for the following strategies:

- (I) the 60-run experiment for the skew models (5.7), from which loss statistics are predicted indirectly, and
- (II) the (40×5) -run, crossed array experiment for modeling the loss statistics directly via Eq. (5.10).

We also consider a hybrid strategy:

- (III) the (40×5) experiment from strategy II but modeling the skews to predict the loss statistics as in strategy I.

Table 5.3 gives results for the loss statistics L_{sq} . Listed there are the best-three sets of transistor dimensions w over the 3^6 grid $\{-1,0,1\}^6$ as predicted by each of these three strategies. The last column gives the true L_{sq} 's from confirmatory experiments. Similarly, Table 5.4 presents the results for the loss statistic L_{wc} . Key features of these results are:

- The confirmatory experiments indicate strategy I predicts the skews accurately enough to give predictions very close to the actual losses.

Table 5.3 Predicted best three circuit designs for the squared-error loss statistic.

| Experiment | Modeling | Best w on 3^6 Cube | L_{sq} | |
|-------------|---------------------|------------------------|-----------|--------|
| | | | Predicted | Actual |
| 60 runs | skews via (5.7) | 0 0 -1 1 1 1 | -0.90 | -0.90 |
| | | 1 1 -1 1 1 1 | -0.76 | -0.77 |
| | | -1 -1 -1 1 1 1 | -0.68 | -0.72 |
| 40 × 5 runs | L_{sq} via (5.10) | -1 1 -1 1 0 1 | -0.38 | -0.50 |
| | | -1 0 -1 1 0 1 | -0.31 | -0.61 |
| | | -1 1 -1 1 -1 1 | -0.30 | -0.41 |
| 40 × 5 runs | skews via (5.7) | 0 0 -1 1 1 1 | -0.89 | -0.90 |
| | | 1 1 -1 1 1 1 | -0.76 | -0.77 |
| | | -1 -1 -1 1 1 1 | -0.69 | -0.72 |

Table 5.4 Predicted best three circuit designs for the worst-case loss statistic.

| Experiment | Modeling | Best w on 3^6 Cube | L_{wc} | |
|-------------|---------------------|------------------------|-----------|--------|
| | | | Predicted | Actual |
| 60 runs | skews via (5.7) | 0 0 -1 1 1 1 | 0.50 | 0.53 |
| | | -1 -1 -1 1 1 1 | 0.56 | 0.51 |
| | | 1 1 -1 1 1 1 | 0.63 | 0.66 |
| 40 × 5 runs | L_{wc} via (5.10) | -1 0 -1 0 0 1 | 0.93 | 1.07 |
| | | -1 0 -1 0 1 1 | 1.05 | 1.17 |
| | | -1 1 -1 0 0 1 | 1.06 | 1.36 |
| 40 × 5 runs | skews via (5.7) | 0 0 -1 1 1 1 | 0.50 | 0.53 |
| | | -1 -1 -1 1 1 1 | 0.52 | 0.51 |
| | | 1 1 -1 1 1 1 | 0.63 | 0.66 |

- Strategy I gives more reliable predictions and superior circuit designs (small actual losses) than strategy II. These differences are of importance; for example, a reduction of the worst-case skews from 1.07 ns to 0.53 ns is of practical significance.
- Strategies I and III give virtually identical results. This shows that a well-chosen, small experiment can be very adequate.
- Comparing the results for all three strategies indicates the superiority of modeling the skews rather than modeling the loss statistics directly.
- Fitting additive models (ignoring interactions) to the loss statistics from a (40×5) run, crossed-array experiment and optimizing the level for each transistor width separately produces even worse predictions. For example, the predicted best squared error is -2.39, but the actual loss computed from a confirmation experiment is -0.58.
- Strategy I gives the same best-three circuit designs for the two loss statistics. This is some indication that even better performance might be obtained by another experiment by changing the ranges of the last four transistors widths. This is borne out by the results for minimizing the predicted loss statistics over the continuous region $[-1,1]^6$ rather than the discrete region $\{-1,0,1\}^6$. Such optimization of the predictor L_{sq} from strategy I leads to $w = \{-0.07, 0.10, -1.00, 1.00, 1.00, 1.00\}$, so that the last four optimal widths are on the boundary. The implication is going beyond the boundary could lead to further improvement.

5.4. Discussion

One reason why our proposed method gives reliable predictions with few observations here is that the skews admit simple models. Moreover, engineering understanding of the underlying circuit performance facilitates model identification. Another, more general advantage of modeling the underlying performance rather than a loss statistic is that collapsing the data to a loss statistic could hide important relationships in the data.

Clearly, more empirical experience is required to determine the general usefulness of the proposed strategy. In another application, a sense-amplifier circuit [47], the proposed method with 60 observations gives more accurate predictions of the losses than modeling the loss statistics with a crossed array experiment of 200 runs. Again, about two-thirds of the observations are saved. In this example there is little difference in the actual performances of the optimized circuit designs.

We used a computer-aided statistical design package to automatically generate the experimental designs. This kind of tool avoids many of the complications often experienced when using catalogued experimental designs. For example, the user can concentrate on the model without worrying about matching the desired interactions to the aliasing structure. We believe that the widespread adoption of these tools and increased attention to modeling rather than combinatorics would encourage the experimentation needed to improve quality. Similarly, model fitting and the minimization of predictions over a grid are straightforward using, for example, SAS [29].

These methods can be extended to physical experiments with random error. In such experiments noise (process) variability is due to unmodeled sources (measurement error, omit-

ted variables, etc.) as well as the manipulation of the noise variables. If the unmodeled sources are unimportant or lead to a noise component with constant variance, there is little technical difficulty in extending our methods. Nonconstant variance requires further study, however.

CHAPTER 6.

CONCLUSIONS

In this dissertation we have presented a new circuit performance modeling approach to statistical circuit design. It is demonstrated that fitted models of the circuit performances, obtained from a statistically designed experiment incurring a small number of circuit simulator runs, can be used as accurate and computationally inexpensive substitutes for the circuit simulator for statistical design and analysis. This modeling approach has been applied to parametric yield prediction [2, 3], yield optimization [7, 8], and achievement of off-line quality control [47, 48]. These algorithms have been implemented into the iEDISON design package [37].

In Chapter 2 we reviewed several methods to represent the distribution of the MOSFET device parameters. It is concluded that the conventional best- and worst-case analyses approach to estimate the range of the device parameter variations is inadequate. An alternative method that uses four critical MOSFET parameters to represent the device parameter distribution is presented. In this method, the other device parameters are treated as functions of the critical parameters. The critical parameter approach is used in our yield prediction and optimization examples in Chapters 3 and 4.

In Chapter 3 we introduced a new method for circuit performance modeling and parametric yield prediction. We modeled the circuit performances by computationally inexpensive approximating functions of the critical parameters. An experimental design that takes replicated observations and a statistical F-test procedure are used to assess the adequacy of

the four critical parameters and the performance model. It is found that sometimes there are more than four critical parameters; engineering knowledge has been used to identify additional critical parameters. A systematic method for critical parameter selection may lead to more reliable results.

The examples in Chapter 3 showed that the circuit performance and parametric yield of many digital circuits can be modeled accurately from about 10 runs of the circuit simulator, if the critical parameters are sufficient. Usually a circuit performance, such as delay and power, can be well approximated by a linear equation with four or five critical parameters, thus, the small number of runs.

The statistical modeling of analog circuit performances is more complicated. Analog circuit performances, such as gain and bandwidth, usually depend strongly on the operation region of the MOSFET transistors (off, linear, or saturation region). If the statistical variations cause a MOSFET to operate outside the region specified by the circuit designer, there may be a dramatic shift in the performance. A linear or quadratic approximation to the performance function may be inadequate. As a result, statistical modeling is not automatic, and engineering knowledge is often needed to identify and fit the model. With careful modeling, however, accurate approximations to the performances can be developed with relatively few circuit simulations.

In Chapter 4 we presented a new parametric yield optimization method. Yield gradient methods for parametric yield maximization usually compute the yield gradient from many circuit simulator runs; this often leads to a large number of runs and hide important relationships in the data. Our method divides yield optimization into two separate steps: (i) Model the circuit performance by an approximating function of the inputs to the circuit simulator;

computationally inexpensive models, optimize the yield with respect to the designable circuit parameters using a proven optimization technique. This approach optimizes the parametric yield with about 100 runs and avoids the complicated yield gradient computation.

In Chapter 5 we applied the circuit performance modeling method to achieve off-line quality control. The Taguchi method for off-line is reviewed. Taguchi's method models the loss statistic of a product, and typically requires a large number of runs. Our method uses fitted models of the circuit performances to predict Taguchi's loss statistic and avoids the nested Taguchi experiments. We showed by example that Taguchi's design objectives can be met by our method with about one-third of the runs.

Clearly, more experience is needed to assess the circuit performance modeling approach. There is substantial industrial interest in this design methodology, and we expect our methods to be tested on larger circuits in the future.

We now present some topics for future research:

- Statistical parameter extraction. The extraction of the device parameters from measured I-V characteristics is usually formulated as curve fitting by non-linear least squares. A more rigorous extraction methodology that states the confidence limits on the parameter estimates is needed.
- Screening of the critical parameters. In this research the four critical MOSFET parameters in [1] are used. These four parameters are found to be inadequate in some situations, however. The more rigorous screening techniques summarized in [19] may lead to better results.

- Mismatch in the device parameters. The device parameter variations within a die are assumed to be insignificant in our research. In high performance analog circuits, intra-die parameter mismatch may lead to significant degradation of the circuit performances. A new methodology to design and analyze circuits with parameter mismatch is desirable.

APPENDIX A.

THE DESIGN OF EXPERIMENT

A.1. Introduction

For a given physical system (circuit), denote the inputs to the system by \mathbf{X} and the response by y . The relation between the response and the inputs is

$$y = h(\mathbf{X}) + \epsilon \quad (\text{A.1})$$

where $h(\mathbf{X})$ is a systematic function of \mathbf{X} , and ϵ is a random error due to variations in the experimental conditions. Typically, ϵ is assumed to take a Gaussian distribution with zero mean and variance σ^2 .

The problem is to

- (i) Identify a model of the system response.
- (ii) Select inputs $\mathbf{X}_1, \dots, \mathbf{X}_N$ to run the experiment, such that approximation $\hat{y}(\mathbf{X})$, obtained from fitting $y(\mathbf{X}_1), \dots, y(\mathbf{X}_N)$ to the assumed model, predicts the response accurately (experimental design problem).

In a physical experiment that has randomness, the fitted model \hat{y} is subjected to two sources of errors:

- The error due to uncertainties in the observations (sampling, or variance error).
- The error due to the systematic departure of the model from the actual response function (lack of fit, or bias error).

Figure A.1 shows an extreme case in which the model and the actual response $h(X)$ have the same form. Due to the random error in the observations, the fitted model is different from $h(X)$ (variance error). For instance, sampling uncertainties lead to distinct fitted models, $\hat{y}_I(X)$ and $\hat{y}_{II}(X)$, in Figure A.1. As the sample size increases, the variance error will decrease to zero.

Figure A.2 shows the other extreme case in which there is no random error, e.g., an experiment conducted on a computer. In this case, least square regression becomes curve fitting. The error in the fitted model $\hat{y}(X)$ will not decrease to zero, even for a large sample size, due to the systematic difference between the model and $h(X)$.

A.2. Integrated Mean-squared Error Criterion

Box and Draper [49] introduced the integrated mean-squared error (IMSE) criterion for the design of experiment. For a given X , the expected squared error of a fitted model is

$$MSE [\hat{y}(X)] = E [\hat{y}(X) - h(X)]^2. \quad (A.2)$$

Integration of Eq. (A.2) over the experimental region R gives

$$\Omega_R \int_R E [\hat{y}(X) - h(X)]^2 dX \quad (A.3)$$

where $1/\Omega_R = \int_R dX$. Normalization of Eq. (A.3) with respect to the number of design points N and the error variance σ^2 , yields the *integrated mean-squared error* (IMSE)

$$J = \frac{N_{OBS} \Omega_R}{\sigma^2} \int_R E [\hat{y}(X) - h(X)]^2 dX. \quad (A.4)$$

The integrated mean-squared error can be decomposed into two components

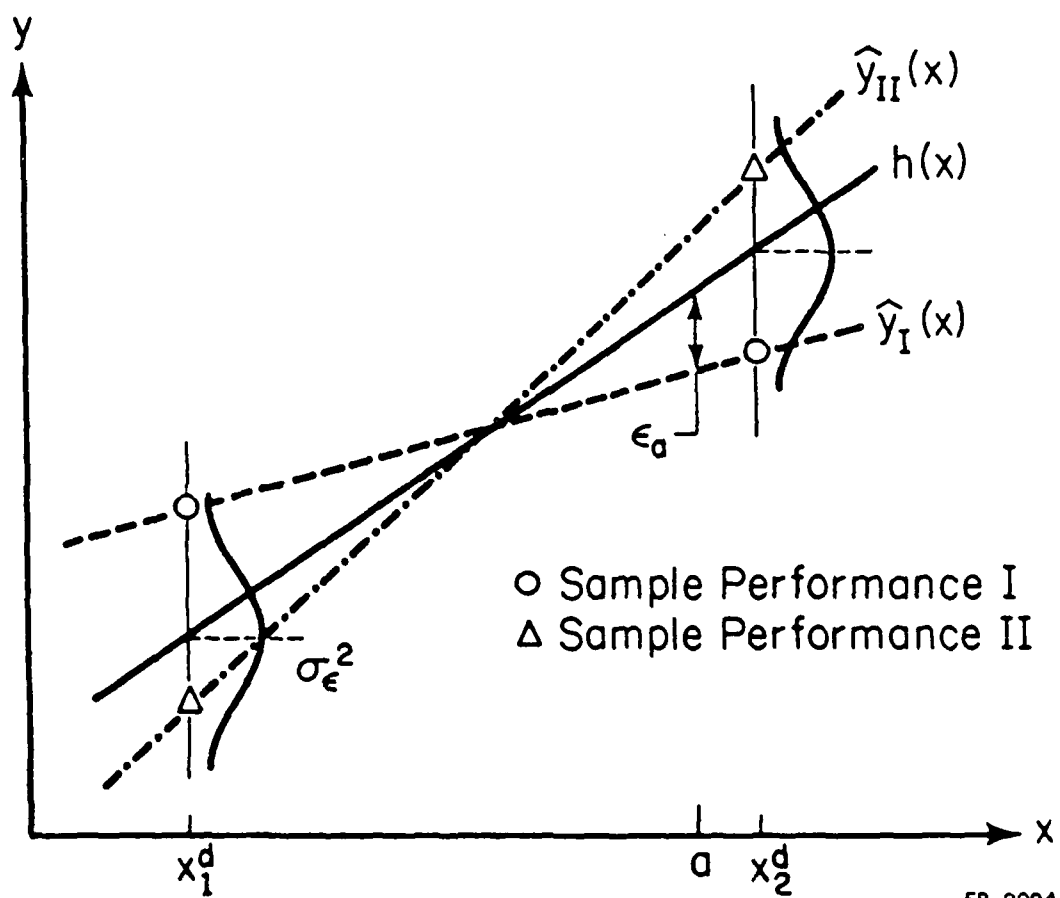
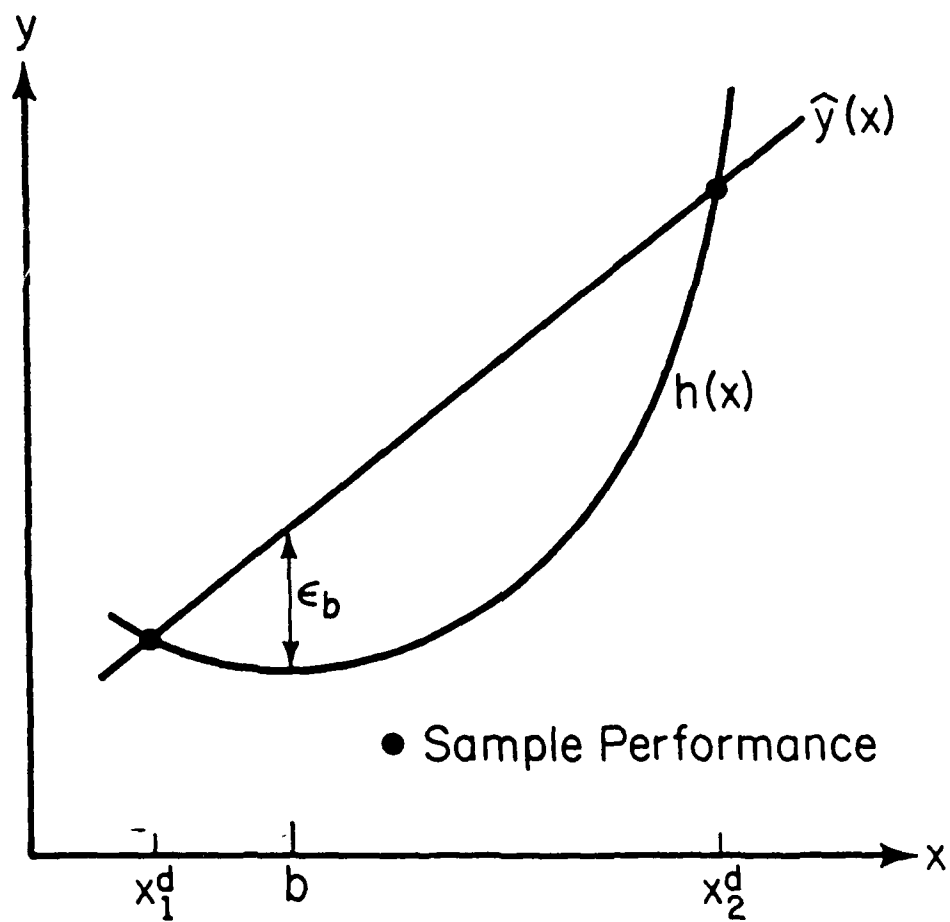


Figure A.1 Model fitting when the variance error is dominant.



FP-9095

Figure A.2 Model fitting when the bias error is dominant.

($J = V + B$), the *average variance*

$$V = \frac{N_{OBS} \Omega_R}{\sigma^2} \int_R \text{Var}(\hat{y}(X)) dX, \quad (\text{A.5})$$

and the *average squared bias*

$$B = \frac{N_{OBS} \Omega_R}{\sigma^2} \int_R [E\hat{y}(X) - h(X)]^2 dX. \quad (\text{A.6})$$

An important point is that whereas V depends only on the location of the design points X_1, \dots, X_N , B depends on both the design points and the actual response function $h(X)$.

To fit an accurate model, the experimental design $D = \{X_1, \dots, X_N\}$ should be chosen such that J is minimized. Welch [50] shows that if the systematic departure of the model from the actual response surface, $|h(X) - E\hat{y}(X)|$ is bounded by Δ for all the X in the region R , then the integrated mean-squared error is

$$J \leq J'(D, \frac{\Delta}{\sigma}) = V(D) + \frac{N\Delta^2}{\sigma^2} B'(D), \quad (\text{A.7})$$

where B' depends only on the experimental design D , and not on the actual response function $h(X)$. This fact provides a guideline for selecting an optimal design that minimizes J' . Note that J' depends on the relative magnitude of the systematic departure and the variance of the random error, $\frac{\Delta}{\sigma}$.

An "excursion" algorithm to select the design D that minimizes J' is presented in [51]. The design of experiment is formulated as a combinatorial optimization problem with cost function J' . These design strategies [50] and [51] are implemented in the program ACED (Algorithms for the Construction of Experimental Designs) program [28].

A.3. Example of Design Point Selection with ACED

Suppose for simplicity that there are only two input parameters x_1 and x_2 . We represent the ranges of x_1 and x_2 by the levels $\{-1,0,1\}$, and the region R by $3^2 = 9$ grid points, as shown in Figure A.3.

The assumed model is

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon \quad (\text{A.8})$$

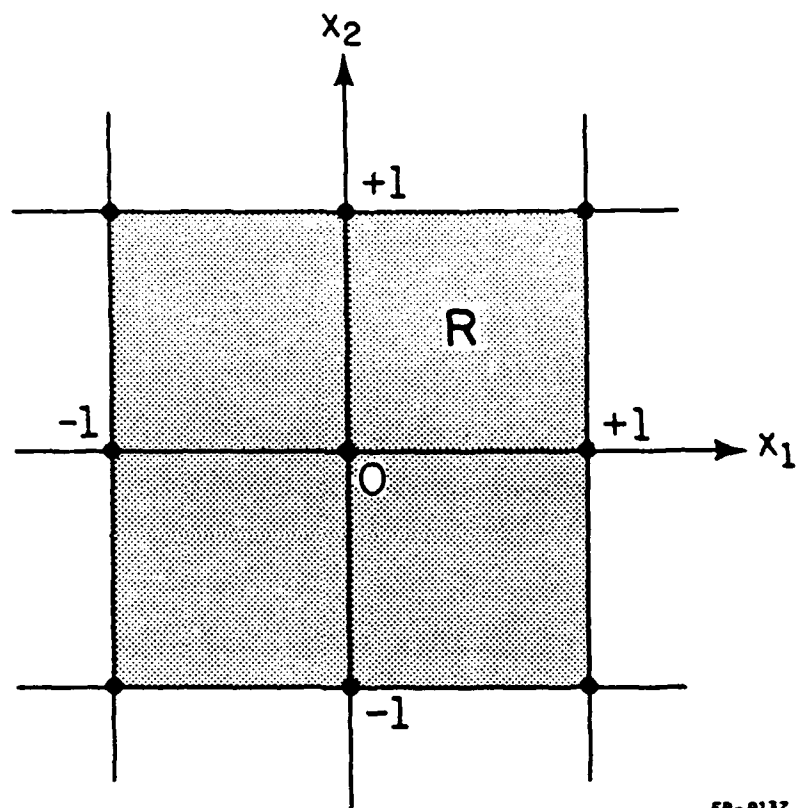
where β_0 , β_1 , and β_2 are the unknown constants, and ε is a random error. The average mean-squared error (am) criterion in ACED is used for the design of experiments. Table A.1 lists the experimental designs that minimize the average mean-squared error for five variance-bias ratios ($\frac{\Delta}{\sigma} = 0, 0.31, 0.55, 1.25$, and ∞), along with their properties. Each design $D^*(\frac{\Delta}{\sigma})$ is denoted in Column 2 by the number of runs at a grid point. For example, the design $D^*(0)$ takes three runs at $(x_1, x_2) = (-1,1)$, and two runs at each of the other three corners of R .

A.3.1. All-variance design

In this case $\frac{\Delta}{\sigma} = 0$, the variance error V is dominant. ACED designs an experiment $D^*(\frac{\Delta}{\sigma} = 0)$ that minimizes V and distributes the points as equally as possible to the corner of the design region.

A.3.2. All-bias design

If there is no random error in the experiment, $\frac{\Delta}{\sigma} = \infty$. ACED designs an experiment



FP-9137

Figure A.3 Two-dimensional design region of 9 grid points.

Table A.1 J' -optimal designs and properties ($N_{OBS} = 9$)

| $\frac{\Delta}{\sigma}$ | $D^*(\frac{\Delta}{\sigma})$ | V | B | $e(D, 0)$ | $e(D, \infty)$ | $e_{\min}(D)$ |
|-------------------------|------------------------------|------|-------|-----------|----------------|---------------|
| 0 | 3 0 2 0 0 0 2 0 2 | 2.39 | 0.922 | 100 | 72.3 | 72.3 |
| 0.31 | 2 1 2 0 0 0 2 0 2 | 2.44 | 0.858 | 98.0 | 77.7 | 77.7 |
| 0.55 | 2 1 1 1 0 1 1 1 1 | 2.64 | 0.750 | 90.5 | 88.9 | 88.9 |
| 1.25 | 1 1 1 1 1 1 1 1 1 | 3.00 | 0.667 | 79.7 | 100.0 | 79.7 |
| ∞ | 1 1 1 1 1 1 1 1 1 | 3.00 | 0.667 | 79.7 | 100.0 | 79.7 |

$D^* (\frac{\Delta}{\sigma} = \infty)$ that minimizes the bias component of the integrated mean-squared error and distributes the points evenly across the design region.

A.3.3. Selection of a robust design

When the relative importance of the variance and bias errors is unknown, ACED searches for a "robust" design good over a wide range of $\frac{\Delta}{\sigma}$. We define the percent efficiency of a design D relative to an optimized design D^* for $\frac{\Delta}{\sigma}$ as

$$e(D, \frac{\Delta}{\sigma}) = \frac{J'(D^*)}{J'(D)} * 100\%. \quad (\text{A.9})$$

For instance, $e(D, 0)$ measures the efficiency of D relative to an optimal all-variance design, $D^* (\frac{\Delta}{\sigma} = 0)$, and $e(D, \infty)$ measures the efficiency of D relative to an optimal all-bias design $D^* (\frac{\Delta}{\sigma} = \infty)$.

A robust design D should have a large efficiency over a large range of variance and bias error ratios. This is measured by

$$e_{\min}(D) = \min e(D, \frac{\Delta}{\sigma}), \quad 0 \leq \frac{\Delta}{\sigma} < \infty. \quad (\text{A.10})$$

Furthermore, it can be shown that e_{\min} is the efficiency with respect to the all-variance or all-bias designs, $e(D, 0)$ or $e(D, \infty)$ [50]. Therefore, a robust design D should maximize

$$e_{\min}(D) = \min [e(D, 0), e(D, \infty)]. \quad (\text{A.11})$$

The minimum efficiency of the J' -optimal design at various values of $\frac{\Delta}{\sigma}$ is shown in Column

5. The ACED program chooses the D^* (0.55) as the most robust design because its minimum efficiency of 88.9% is larger than the minimum efficiency of the other designs.

A.4. Design and Analysis of Computer Experiments

The classical experimental designs, associated with Box and co-workers [24,26], are invented for physical experiments that have random errors. However, in circuit simulation, the computer code always produces the same outputs for the same inputs.

As a result, there are two ways to design and analyze a computer-simulation experiment:

- (1) In addition to the experimental factors X , vary the other inputs to the computer code according to their distribution. In the presence of random errors, classical designs can be used for the experiment.
- (2) With the exception of the experimental factors X , fix all the inputs to the computer code at their nominal values. In the absence of random errors, an all-bias design from ACED can be used for the experiment.

The design and analysis of deterministic computer-simulation experiments are discussed in [35]:

- The absence of random error allows the complexity of the actual response surface to emerge.
- The adequacy of the fitted response surface model is determined solely by the systematic departure of the model from the actual response.

- The adequacy of the fitted response surface model is determined solely by the systematic departure of the model from the actual response.
- There is no obvious statistical basis for estimating the uncertainty.
- Conventional notations of experimental unit, blocking, replication, and randomization are irrelevant.

As a consequence, it is unclear if the current methods [24] for the design and analysis of physical experiments are ideal for computer simulations. However, Sacks, Welch, Mitchell, and Wynn [35] assert that

- The selection of inputs at which to run a computer code is still an experimental design problem.
- Statistical principles and attitudes to data analysis are helpful however the data are generated.
- There is uncertainty associated with predictions from the fitted models, and the quantification of uncertainty is a statistical analysis problem.

A.5. Latin Hypercube Design

McKay, Beckman, and Conover [25] were among the first to explicitly consider the statistical analysis of deterministic computer codes. They introduced Latin hypercube sampling, and advocated it as an alternative to simple random sampling in Monte Carlo studies (see Section 2.1). The objective of Latin hypercube sampling is to determine, for a known input distribution, the distributions of the outputs.

- STEP 1: Sample the range of each input variable x_k , $k = 1, \dots, p$, into N (evenly-spaced) points, and denote the samples by x_{kj} , $j = 1, \dots, N$.
- STEP 2: The samples x_{kj} , $j = 1, \dots, N$, will form the x_k component, $k = 1, \dots, p$, in X_i , $i = 1, \dots, N$. The components of the various x_k 's are matched at random to spread the samples evenly in the input space.

Steck, Iman, and Dahlgren [52] and Atwood [53] advocated using the Latin hypercube sampling method for the design of computer experiments. The study in [52] shows Latin hypercube designs (samples) are superior to classical fractional factorials designs [26] for fitting response surfaces in several computer experiments.

Compared to designs from the average mean-squared error criterion in ACED, the Latin hypercube designs are much cheaper and they can easily handle problems of high dimensions. For a small experiment, however, an all-bias design from ACED may outperform a Latin hypercube design that has the same number of runs.

A FORTRAN program to generate Latin hypercube designs is given below:

```

program latin

c      n      = number of independent factors
c      nobs   = number of runs in the experiment
c      lhs(nobs,m) = matrix of latin hypercube design

parameter (m=5, nobs=12)

integer iseed
integer lhs(nobs, m), iper(nobs)
real    x(nobs), delta

integer i,j

```

```

        external rper, rset

c...Discretize [-1,1] interval into nobs intervals of length delta
        delta = 2.0 / (nobs-1)

        do 10 i= 0, nobs-1, 1
10      x(i+1) = -1.0 + i * delta

c...Setup with IMSL, initialize random number generator

        iseed = 1234567
        call rset(iseed)

c...Random permutation for column i=1,...,m

        do 20 j=1,m,1
            call rper(nobs, iper)
            do 30 i=1,nobs,1
                lhs(i,j) = iper(i)
30          continue
20        continue

c...Print out Latin Hypercube design

        write (6,800) m, nobs
        do 50 i=1,nobs,1
            write (6,900) i, (x(lhs(i,j))), j=1,m,1)
50        continue
        stop

800      format(' Latin hypercube design of ', i5, ' factors with', i5,
1         ' observations'//)
900      format(i2,2x,100F10.4)
        end

```


APPENDIX B.

MODEL ASSESSMENT

B.1. A Statistical F-test Procedure for Model Assessment

For a given regression model, denote the least square prediction of the i 'th data point, $y(x_i)$, by $\hat{y}(x_i)$ for $i = 1, \dots, N$. The procedure in [30] considers a fitted model to be adequate if the following conditions are satisfied:

- (a) The variation in $y(x_i)$, $i = 1, \dots, N$, "explained" by $\hat{y}(x)$ is substantially larger than the error in the regression.
- (b) The lack of fit (LOF) of data to the model is insignificant.

The sum of squares variations explained by the model is

$$SS_{Model} = \sum_{i=1}^N [\hat{y}(x_i) - y_{avg}]^2, \quad (B.1)$$

where y_{avg} is the average of $y(x_i)$, $i = 1, \dots, N$.

The error sum of squares is

$$SS_{Error} = \sum_{i=1}^N [\hat{y}(x_i) - y(x_i)]^2. \quad (B.2)$$

The normalized regression sum of squares is

$$MS_{Model} = \frac{SS_{Model}}{M} \quad (B.3)$$

and the normalized sum of squares of the errors is

$$MS_{Error} = \frac{SS_{Model}}{N - M - 1}. \quad (B.4)$$

A statistical F-test is introduced in [30] to test condition (a). The goal is to determine whether the expected variation due to the model $E[MS_{Model}]$ is at least γ_o^2 times larger than the expected error, $E[MS_{Error}]$ (typically $\gamma_o^2 = 2$ to 4).

The F-statistic of the regression

$$F_1 = \frac{MS_{Model}}{MS_{Error}} \quad (B.5)$$

is compared with the corresponding critical F-value, $F_{1,cr}$. The critical $F_{1,cr}$ is the $1 - \alpha$ percentile point of a noncentral F-distribution with M and $N - M - 1$ degrees of freedom, and the noncentrality parameter γ_o^2 . If F_1 is larger than $F_{1,cr}$, condition (a) is considered satisfied.

To assess the lack of fit in the model, we partition the sum of squares error into "pure error" (PE) and "lack of fit" (LOF) components. The "pure error" is due to random variations only. If there are k replicated runs at $x, y_{(1)}, \dots, y_{(k)}$, the "pure error" sum of squares is

$$\sum_{i=1}^k [y_{(i)}(x) - \bar{y}]^2, \quad (B.6)$$

where \bar{y} is the average response at x . The total pure error sum of squares, SS_{PE} , is the sum of the individual "pure error" terms over all x . We denote the degrees of freedom due to pure error by K.

The sum of squares due to the lack of fit is

$$SS_{LOF} = SS_{Error} - SS_{PE}. \quad (B.7)$$

It has $N - M - K - 1$ degrees of freedom. The normalized sum of squares due to pure error and lack of fit is denoted by MS_{PE} and MS_{LOF} , respectively.

The F-statistic

$$F_2 = \frac{MS_{LOF}}{MS_{PE}} \quad (B.8)$$

estimates the importance of the lack of fit. We compare F_2 with $F_{2,cr}$. The critical value $F_{2,cr}$ is the $1 - \alpha$ percentile point of a central F-distribution with K and $N - M - K - 1$ degrees of freedom. If F_2 is less than $F_{2,cr}$, condition (b) is considered satisfied. The relations between the sum of squares and F-statistics are summarized in an analysis of variance (ANOVA) table (Table B.1).

The F-tests have three possible outcomes:

- Case 1. If $F_1 > F_{1,cr}$, condition (a) is satisfied. The fitted model $\hat{y}(X)$ is considered adequate.
- Case 2. If $F_1 < F_{1,cr}$ and $F_2 < F_{2,cr}$, condition (a) is violated but the lack of fit is insignificant. Hence the fitted model is declared inadequate. The model inadequacy may be caused by either the small sample size or the effects of parameters that are not included in the model.
- Case 3. If $F_1 < F_{1,cr}$ and $F_2 > F_{2,cr}$, condition (a) is violated and the lack of fit is significant. The fitted model is considered inadequate, and a more complex model may be needed to represent the response surface.

Table B.1 ANOVA table for the regression model.

| Source | SS | df | MS | F-ratio | Test Statistic |
|-------------|--------------|---------|---|---------------------------------------|----------------|
| Model | SS_{Model} | M | $MS_{Model} = \frac{SS_{Model}}{N}$ | $F_1 = \frac{MS_{Model}}{MS_{Error}}$ | $F_{1,cr}$ |
| Error | SS_{Error} | N-M-1 | $MS_{Error} = \frac{SS_{Error}}{N-M-1}$ | | |
| Lack of Fit | SS_{LOF} | N-M-K-1 | $MS_{LOF} = \frac{SS_{LOF}}{N-M-K-1}$ | $F_2 = \frac{MS_{LOF}}{MS_{PE}}$ | $F_{2,cr}$ |
| Pure Error | SS_{PE} | K | $MS_{PE} = \frac{SS_{PE}}{K}$ | | |

B.2. Assessing the Goodness of Fit

For a given regression model, denote the least square prediction of the i 'th data point, $y(x_i)$, by $\hat{y}(x_i)$ for $i = 1, \dots, N$. Then the squared error of prediction is $[\hat{y}(x_i) - y(x_i)]^2$. The R^2 -statistic relates the total of these squared errors to the variability in the data:

$$R^2 = 1 - \frac{\sum_{i=1}^N [\hat{y}(x_i) - y(x_i)]^2}{\sum_{i=1}^N [y(x_i) - \bar{y}]^2} \quad (\text{B.9})$$

where \bar{y} is the mean of $y(x_1), \dots, y(x_N)$. It can be shown that $0 \leq R^2 \leq 1$, with larger values suggesting better agreement between the predictions and the model.

One problem with R^2 is that it tends to overestimate the predictive power of a regression model, because the same data are used to fit and to test the regression. A more stringent test is based on predicting $y(x_i)$ by $\hat{y}_{-i}(x_i)$, where $\hat{y}_{-i}(x_i)$ is the least squares prediction based on all the data *except* the i 'th case. This leads to

$$R^2_{PRS} = 1 - \frac{\sum_{i=1}^N [\hat{y}_{-i}(x_i) - y(x_i)]^2}{\sum_{i=1}^N [y(x_i) - \bar{y}]^2}. \quad (\text{B.10})$$

Again, the R^2_{PRS} value of 1 indicates perfect agreement between the predictor and the data, but R^2_{PRS} can be much less than R^2 (even less than 0!), indicating possibly poor predictive capability and lack of fit.

REFERENCES

- [1] P. Cox, P. Yang, S. S. Manhant-Shetti, and P. Chatterjee, "Statistical modeling for efficient parametric yield estimation of MOS VLSI circuits," *IEEE J. Solid State Circuits*, vol. SC-20, no. 1, pp. 391-398, Feb. 1985.
- [2] T. K. Yu, S. M. Kang, I. N. Hajj, and T. N. Trick, "Statistical modeling of VLSI circuit performances," *IEEE Int. Conf. Computer-Aided Design*, Santa Clara, CA, Nov. 1986.
- [3] T. K. Yu, S. M. Kang, I. N. Hajj, and T. N. Trick, "Statistical performance modeling and parametric yield estimation of MOS VLSI," *IEEE Trans. Computer-Aided Design*, vol. CAD-6, no. 6, pp. 1013-1022, Nov. 1987.
- [4] W. Maly, A. J. Strojwas, and S. W. Director, "VLSI yield prediction and estimation: a unified framework," *IEEE Trans. Computer-Aided Design*, vol. CAD-5, pp. 114-130, Jan. 1986.
- [5] V. Visvanathan, "Variational analysis of integrated circuits," *IEEE Int. Conf. Computer-Aided Design*, pp. 228-231, Santa Clara, CA, Nov. 1986.
- [6] M. A. Styblinski and L. J. Opalski, "Algorithms and software tools for IC yield optimization based on fundamental fabrication parameters," *IEEE Trans. Computer-Aided Design*, vol. CAD-5, no. 1, pp. 79-89, Jan. 1986.
- [7] T. K. Yu, S. M. Kang, J. Sacks, and W. J. Welch, "An efficient method for parametric yield optimization of MOS integrated circuits," *IEEE Int. Conf. Computer-Aided Design*, Santa Clara, CA, Nov. 1989.
- [8] T. K. Yu, S. M. Kang, J. Sacks, and W. J. Welch, "Parametric yield optimization of MOS integrated circuits by statistical modeling of circuit performances," *submitted to IEEE Trans. Computer-Aided Design*, July 1989.
- [9] G. Taguchi, *Introduction to Quality Engineering*, Asian Productivity Organization, Tokyo, 1986.
- [10] S. R. Nassif, A. J. Strojwas, and S. W. Director, "A method for worst-case analysis of integrated circuits," *IEEE Trans. Computer-Aided Design*, vol. CAD-5, no. 1, pp. 104-113, Jan. 1986.
- [11] P. Yang and P. Chatterjee, "Statistical modelling of small geometry MOSFETs," *IEEE Int. Electron Device Meeting*, pp. 286-289, Dec. 1982.
- [12] N. Herr and J. J. Barnes, "Statistical circuit simulation modeling of CMOS VLSI," *IEEE Trans. Computer-Aided Design*, vol. CAD-5, no. 1, pp. 15-22, Jan. 1986.

- [13] S. R. Nassif, A. J. Strojwas, and S. W. Director, "FABRICS II - A statistically based IC fabrication process simulator," *IEEE Trans. Computer-Aided Design*, vol. CAD-3, no. 1, pp. 40-46, Jan. 1984.
- [14] S. Selberherr, A. Schutz, and H. Potzl, "Investigation of parameter sensitivity of short channel MOSFETs," *Solid-State Electron.*, vol. 25, no. 2, pp. 85-90, 1982.
- [15] P. Yang and P. K. Chatterjee, "SPICE modeling for small geometry MOSFET circuits," *IEEE Trans. Computer-Aided Design*, vol. CAD-1, pp. 169-182, 1982.
- [16] S. Liu and K. Singhal, "A statistical model for MOSFETs," *IEEE Int. Conf. Computer-Aided Design*, pp. 78-80, Santa Clara, CA, Nov. 1985.
- [17] N. R. Draper and H. Smith, *Applied Regression Analysis*, 2nd ed., New York: Wiley, 1981.
- [18] J. P. Spoto, W. T. Coston, and C. P. Hernandez, "Statistical integrated circuit design and characterization," *IEEE Trans. Computer-Aided Design*, vol. CAD-5, no. 1, pp. 90-103, Jan. 1986.
- [19] D. J. Downing, R. H. Gardner, and F. O. Hoffman, "An examination of response-surface methodologies for uncertainty analysis in assessment models," *Technometrics*, vol. 27, no. 2, pp. 151-163, May 1985.
- [20] R. L. Iman, J. C. Helton, and J. E. Campbell, "An approach to sensitivity analysis of computer models: Part I - introduction, input variable selection and preliminary variable assessment," *J. Quality Technol.*, vol. 13, no. 3, pp. 174-183, July 1981.
- [21] R. L. Iman, J. C. Helton, and J. E. Campbell, "An approach to sensitivity analysis of computer models: Part II - ranking of input variables, response surface validation, distribution effect and technique synopsis," *J. Quality Technol.*, vol. 13, no. 4, pp. 232-240, Oct. 1981.
- [22] W. T. Weeks, A. J. Jimenez, G. W. Mahoney, D. Mehta, H. Qassemzadeh, and T. R. Scott, "Algorithms for ASTAP - a network-analysis program," *IEEE Trans. Circuit Theory*, vol. CT-20, pp. 628-634, Nov. 1973.
- [23] D. E. Hocevar, M. R. Lightner, and T. N. Trick, "A study of variance reduction techniques for estimating circuit yields," *IEEE Trans. Computer-Aided Design*, vol. CAD-2, no. 3, pp. 180-192, July 1983.
- [24] G. E. P. Box and N. R. Draper, *Empirical Model-building and Response Surfaces*, New York: Wiley, 1987.
- [25] M. D. McKay, R. J. Beckman, and W. J. Conover, "A comparison of three methods for selecting values of input variables in the analysis of output from a computer code," *Technometrics*, vol. 21, no. 2, pp. 239-245, May 1979.
- [26] G. E. P. Box, W. G. Hunter, and J. S. Hunter, *Statistics for Experimenters*, New York: Wiley, 1978.

- [27] S. M. Kang, "Accurate simulation of power dissipation in VLSI circuits," *IEEE J. Solid State Circuits*, vol. SC-21, no. 5, pp. 889-891, Oct. 1986.
- [28] W. J. Welch, "ACED: algorithms for the construction of experimental designs," *American Statistician*, pp. 205-213, 1985.
- [29] *SAS User's Guide: Basics and Statistics, 5th ed.*, Cary, NC: SAS Institute, 1985.
- [30] R. Suich and G. C. Derringer, "Is the regression model adequate? - one criterion," *Technometrics*, vol. 19, no. 2, pp. 213-216, May 1977.
- [31] R. H. Krambeck, C. M. Lee, and H.-F. S. Law, "High-speed compact circuits with CMOS," *IEEE J. Solid State Circuits*, vol. SC-27, no. 3, pp. 614-619, June 1982.
- [32] K. Singhal and J. F. Pinel, "Statistical design centering and tolerancing using parametric sampling," *IEEE Trans. Circuits Syst.*, vol. CAS-28, pp. 691-702, July 1981.
- [33] D. E. Hocevar, P. F. Cox, and P. Yang, "Parametric yield optimization for MOS circuit blocks," *IEEE Trans. Computer-Aided Design*, vol. CAD-7, no. 6, pp. 645-658, June 1988.
- [34] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes*, New York: Cambridge, 1986.
- [35] J. Sacks, W. J. Welch, T. J. Mitchell, and H. P. Wynn, "Design and analysis of computer experiments," *Statistical Science*, to appear Nov. 1989.
- [36] P. E. Allen and D. A. Holdberg, *CMOS Analog Circuit Design*, New York: Holt, Rinehart, and Winston, 1987.
- [37] T. K. Yu, S. M. Kang, I. N. Hajj, and T. N. Trick, "iEDISON: an interactive statistical design tool for MOS VLSI circuits," *IEEE Int. Conf. Computer-Aided Design*, pp. 20-23, Santa Clara, CA, Nov. 1988.
- [38] D. J. Allstot, "A precision variable-supply CMOS comparator," *IEEE J. Solid State Circuits*, vol. SC-17, no. 6, pp. 1080-1087, Dec. 1982.
- [39] R. N. Kacker, "Off-line quality control, parameter design, and the Taguchi method (with discussions)," *J. Quality Technol.*, vol. 17, no. 4, pp. 176-209, Oct. 1985.
- [40] J. S. Hunter, "Statistical design applied to product design," *J. Quality Technol.*, vol. 17, no. 4, pp. 210-221, Oct. 1985.
- [41] R. N. Kacker and A. C. Shoemaker, "Robust design: a cost effective method for improving manufacturing processes," *AT&T Technical J.*, vol. 65, pp. 39-50, March-April 1986.

- [42] M. S. Phadke, "Design optimization case studies," *AT&T Technical J.*, vol. 65, pp. 51-68, March-April 1986.
- [43] R. V. Leon, A. C. Shoemaker, and R. N. Kackar, "Performance measures independent of adjustment (with discussions)," *Technometrics*, vol. 29, no. 3, pp. 253-285, Aug. 1987.
- [44] G. E. P. Box, "Signal-to-noise ratios, performance criteria, and transformations (with discussions)," *Technometrics*, vol. 30, no. 1, pp. 1-40, Feb. 1988.
- [45] M. Shoji, "Elimination of process-dependent clock skew in CMOS VLSI," *IEEE J. Solid State Circuits*, vol. SC-21, no. 5, pp. 875-880, Oct. 1986.
- [46] R. G. Easterling, "Discussion of 'Off-line quality control, parameter design and the Taguchi method (with discussions)' by R. N. Kackar," *J. Quality Technol.*, vol. 17, pp. 191-192, Oct. 1985.
- [47] T. K. Yu and S. M. Kang, "Statistical MOS VLSI circuit optimization with non-nested experimental design," *Int. Symp. Circuits Syst.*, pp. 1799-1802, June 1988.
- [48] W. J. Welch, T. K. Yu, S. M. Kang, and J. Sacks, "Computer experiments for quality control by parameter design," *J. Quality Technol.*, to appear Jan. 1990.
- [49] G. E. P. Box and N. R. Draper, "A basis for the selection of a response surface design," *American Statistical Assoc. J.*, vol. 54, pp. 622-654, Sept. 1959.
- [50] W. J. Welch, "A mean squared error criterion for the design of experiments," *Biometrika*, vol. 70, pp. 205-213, 1983.
- [51] W. J. Welch, "Computer-aided design of experiments for response estimation," *Technometrics*, vol. 26, pp. 217-224, Aug. 1984.
- [52] G. P. Steck, R. L. Iman, and D. A. Dahlgren, "Probabilistic analysis of LOCA," *Annual report for FY 1976, SAND76-0535*, pp. 75-82, Sandia Lab., Albuquerque, NM, 1976.
- [53] C. L. Atwood, "Comments on 'Small sample sensitivity analysis techniques for computer models,' by R. L. Iman and W. J. Conover," *Comm. in Statistics - Theor. Meth.*, vol. A9, no. 17, pp. 1847-1851, 1980.

VITA

Tat-Kwan E. Yu [REDACTED] He received his B.S.E. degree in Electrical and Computer Engineering from the University of Michigan, Ann Arbor, Michigan, in 1982. In 1983 he entered the University of Illinois at Urbana-Champaign and received his M.S. degree in Electrical Engineering in 1985. From January 1983 to June 1989 he worked as a Research Assistant at the Coordinated Science Laboratory, University of Illinois, and as a Teaching Assistant in the Department of Electrical and Computer Engineering at the University of Illinois. He has accepted a position as a staff engineer in the Advanced Products Research and Development Laboratory at Motorola Inc., at Austin, Texas. His research interests include the areas of statistical design, quality control, and simulation of integrated circuits.